

REPORT DOCUMENTATION PAGE				Form Approved OMB NO. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA, 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</p>					
1. REPORT DATE (DD-MM-YYYY) 12-10-2010		2. REPORT TYPE Final Report		3. DATES COVERED (From - To) 1-Mar-2005 - 30-Sep-2010	
4. TITLE AND SUBTITLE Multi-Frequency Analysis for Landmine Detection with Forward-Looking Ground Penetrating Radar				5a. CONTRACT NUMBER W911NF-05-1-0069	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 633606	
6. AUTHORS James M. Keller, Dominic K.C. Ho				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES AND ADDRESSES University of Missouri - Columbia Office of Sponsored Programs The Curators of the University of Missouri Columbia, MO 65211 -				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211				10. SPONSOR/MONITOR'S ACRONYM(S) ARO	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) 48343-EV.1	
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.					
14. ABSTRACT The project investigates image processing, sensor fusion and signal processing techniques for the forward-looking ground penetrating radar (FLGPR) explosive detection system equipped with a color or FLIR camera (the Alaric system fielded by NVESD), as well as independent multi-camera systems. Also, in this report period, we are addressing research issues dealing with feature and sensor fusion. We had some partial funding from a Leonard Wood Institute Grant. The ultimate goal is to utilize multiple sensing modalities together with FLGPR to increase					
15. SUBJECT TERMS FLGPR, FLIR, Color imagery, human-in-the-loop, explosive hazard detection, image modality registration, fusion					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON James Keller
a. REPORT UU	b. ABSTRACT UU	c. THIS PAGE UU			19b. TELEPHONE NUMBER 573-882-7339

Report Title

Multi-Frequency Analysis for Landmine Detection with Forward-Looking Ground Penetrating Radar

ABSTRACT

The project investigates image processing, sensor fusion and signal processing techniques for the forward-looking ground penetrating radar (FLGPR) explosive detection system equipped with a color or FLIR camera (the Alaric system fielded by NVESD), as well as independent multi-camera systems. Also, in this report period, we are addressing research issues dealing with feature and sensor fusion. We had some partial funding from a Leonard wood Institute Grant. The ultimate goal is to utilize multiple sensing modalities together with FLGPR to increase IED detection with low false alarm rates. The project objectives are to: Perform image processing for infra-red and color cameras to detect surface laid road-side targets; Investigate advanced target detection approaches for the FLGPR; Develop coordinate mapping technique between EO image sensors and FLGPR data and investigate fusion algorithms; Research and develop approaches for vehicle-based human-in-the-loop cuing of explosive devices using EO sensors; and Examine and process the EO and FLGPR data collected by the U.S. Army and improve algorithm performance through extensive testing.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

K. C. Ho, L. Carin, P. D. Gader and J. N. Wilson, "On using the spectral features from ground penetrating radar for landmine/clutter discrimination," IEEE Trans. Geosci. Remote Sensing, vol. 46, pp. 1177-1191, Apr. 2008..

J. N. Wilson, P. D. Gader, W. H. Lee, H. Frigui and K. C. Ho, A rigorous evaluation of algorithms using ground penetrating radar for landmine detection and discrimination, IEEE Trans. Geosci. Remote Sensing, vol. 45, pp. 2560-2572, Aug. 2007.

Mendez-Vazquez, A., Gader, P., Keller, J., and Chamberlin, K., "Minimum classification error training for Choquet integrals with applications to landmine detection", IEEE Transactions on Fuzzy Systems, Vol. 16, No. 1, 2008, pp. 225-238.

Wang, T., Keller, J., Gader, P., and Sjahputera, O., "Frequency Subband Processing and Feature Analysis of Forward-Looking Ground Penetrating Radar Signals for Land Mine Detection", IEEE Transactions on Geoscience and Remote Sensing, Vol. 45, No. 3, 2007 pp. 718 - 729.

Number of Papers published in peer-reviewed journals: 4.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

Wang, T., Keller, J., Gader, P., and Sjahputera, O., "Land Mine Detection in Forward-Looking Ground Penetrating Radar Using Frequency Subband Processing and Feature Analysis", invited presentation, SIAM Annual Meeting, New Orleans, LA, July, 2005.

Keller, J., Ho, D., Wang, T., Gader, P., Ramachandran, G., Hawkins, C., Elroy, J., Wilson, J., "Research in Forward Looking Landmine and IED Detection", Army Landmine BRTR, Springfield, VA, February, 2006.

Keller, J., Ho, D., "Research in Forward Looking Landmine and IED Detection", Army Landmine BRTR, Springfield, VA, January, 2007.

K. C. Ho, P. D. Gader and J. N. Wilson, Subspace processing on the energy density spectrum for landmine detection using GPR, UXO Forum, Aug. 2007.

J. N. Wilson, K. C. Ho, P. D. Gader, S. Burke, R. Cresci, P. Ngan, Realtime processing algorithm for the AN/PSS14, UXO Forum, Aug. 2007.

Keller, J., Ho, D., "Research in Forward Looking Landmine and IED Detection", Army Landmine BRTR, Springfield, VA, January, 2008.

Keller, J., Ho, D., "Research in Forward Looking Landmine and IED Detection", Army Landmine BRTR, Springfield, VA, January, 2009.

Keller, J., Ho, D., "Research in Forward Looking Landmine and IED Detection", Army Landmine BRTR, Washington, DC, January, 2010.

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Wang, T., Keller, J., Gader, P., and Sjahputera, O., “Land Mine Detection in Forward-Looking Ground Penetrating Radar Using Frequency Subband Processing and Feature Analysis”, invited presentation, SIAM Annual Meeting, New Orleans, LA, July, 2005.

Wang, T., Sjahputera, O., Keller, J., and Gader, P., “Landmine detection using forward-looking GPR with object-tracking”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets X, Orlando, FL, March/April 2005, pp. 1080-1088.

Wang, T., Sjahputera, O., Keller, J., and Gader, P., “Feature analysis for forward-looking landmine detection using GPR”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets X, Orlando, FL, March/April 2005, pp. 1233-1244.

Wang, T., Keller, J., Busch, M., Gader, P., Hawkins, C., McElroy, J., and Ho, K.C., “On The Confidence Level Fusion of IR and Forward-Looking GPR”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XI, Orlando, FL, April 2006, pp. 62172T-1 to 62172T-9.

Busch, M., Keller, J., and Gader, P., “A scale space approach to detect a class of side-attack landmines from SWIR video sequences, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XI, Orlando, FL, April 2006, pp. 621730-1 to 621730-8.

Stone, K., Keller, J., Ho, K.C., Busch, M., Gader, “On the registration of FLGPR and IR data for a forward-looking landmine detection system and its use in eliminating FLGPR false alarms”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XIII, Orlando, FL, March, 2008, pp. 6953 14-1 to 6953 14-11.

K. C. Ho, P. D. Gader, J. N. Wilson, H. Frigui, Subspace processing of ground penetrating radar and its fusion for landmine detection, in Proc. SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets XIII, Orlando, Mar. 2008.

Keller, J. M., Stone, K. E., Ho, K. C., and Popescu, M., “Automatic cuing of human-in-the-loop detection system,” in Proc. SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets XIV, Orlando, Apr. 2009.

T. C. Havens, T. C., Keller, J. M., Ho, K. C., and Stone, K. E., “Sensor-fused detection of explosive hazards,” in Proc. SPIE Conf. Detection and Remediation Technologies for Mines and Minelike Targets XIV, Orlando, Apr. 2009.

Greenwood, G., Blakely, S., Schartman, D., Calhoun, B., Keller, J., Ton, T. Wong, D., and Soumekh, M., “Feature Extraction and Object Recognition in Multi-Modal Forward Looking Imagery”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV, Orlando, FL, April, 2010.

Havens, T., Ho, K.C., Farrell, J., Keller, J., Popescu, M., Ton, T. Wong, D., and Soumekh, M., “Locally-Adaptive Detection Algorithm for Forward-Looking Ground-Penetrating Radar”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV, Orlando, FL, April, 2010.

Popescu, M., Stone, K., Havens, T., Ho, K.C., and Keller, J., “Anomaly Detection in Forward Looking Infrared Imaging Using One Class Classifiers”, ”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV, Orlando, FL, April, 2010.

Stone, K., Keller, J., Popescu, M., Havens, T., and Ho, K.C., Forward Looking Anomaly Detection via Fusion of Infrared and Color Imagery, ”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV, Orlando, FL, April, 2010.

Havens, T., Spain, C., Ho, K.C., Keller, J., Ton, T. Wong, D., and Soumekh, M., “Improved Detection and False Alarm Rejection Using FLGPR and Color Imagery in a Forward-Looking System”, ”, Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV, Orlando, FL, April, 2010.

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

14

Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Number of Manuscripts: 0.00

Patents Submitted

Patents Awarded

Graduate Students

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
Kevin Stone (in 2008)	1.00
Kevin Stone (otherwise)	0.50
Tim Havens	0.50
Chris Spain	0.50
Justin Farrell	0.50
David Lewis	0.50
FTE Equivalent:	3.50
Total Number:	6

Names of Post Doctorates

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
Derek Anderson	1.00
FTE Equivalent:	1.00
Total Number:	1

Names of Faculty Supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>	National Academy Member
James Keller	0.15	No
Dominic Ho	0.15	No
Mihail Popescu	0.10	No
FTE Equivalent:	0.40	
Total Number:	3	

Names of Under Graduate students supported

<u>NAME</u>	<u>PERCENT SUPPORTED</u>
Don Scharzman	0.30
Brad Calhoun	0.30
FTE Equivalent:	0.60
Total Number:	2

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 1.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:..... 1.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:..... 0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):..... 1.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:..... 0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields: 0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Timothy Havens

Total Number: 1

Names of other research staff

NAME

Mark Busch

PERCENT SUPPORTED

1.00 No

FTE Equivalent: 1.00

Total Number: 1

Sub Contractors (DD882)

1 a. Portland State University

1 b. P.O. Box 751

Portland

OR

97207

Sub Contractor Numbers (c): C00016486-1

Patent Clause Number (d-1):

Patent Date (d-2):

Work Description (e): Feature extraction and object class description for video data related to Explosive Hazard detection.

Sub Contract Award Date (f-1): 7/1/2009 12:00:00AM

Sub Contract Est Completion Date(f-2): 8/31/2010 12:00:00AM

Inventions (DD882)

2010 Project Report

Multi-Frequency Analysis for Landmine Detection with Forward-Looking Ground Penetrating Radar

Proposal Number 48343-EV

James Keller and Dominic Ho

Department of Electrical and Computer Engineering
University of Missouri-Columbia

Report Period: September 1, 2009 to August 31, 2010

Objectives

The project investigates image processing, sensor fusion and signal processing techniques for the forward-looking ground penetrating radar (FLGPR) explosive detection system equipped with a color or FLIR camera (the Alaric system fielded by NVESD), as well as independent multi-camera systems. Also, in this report period, we are addressing research issues dealing with feature and sensor fusion. We had some partial funding from a Leonard wood Institute Grant. The ultimate goal is to utilize multiple sensing modalities together with FLGPR to increase IED detection with low false alarm rates. The project objectives are to

- Perform image processing for infra-red and color cameras to detect surface laid road-side targets.
- Investigate advanced target detection approaches for the FLGPR.
- Develop coordinate mapping technique between EO image sensors and FLGPR data and investigate fusion algorithms.
- Research and develop approaches for vehicle-based human-in-the-loop cuing of explosive devices using EO sensors.
- Examine and process the EO and FLGPR data collected by the U.S. Army and improve algorithm performance through extensive testing.

Approach

We have made considerable progress in the past year on a variety of approaches that examine the utility of EO sensors (alone and with fusion), direct detection vs. change detection, the fusion of those formats, new approaches to explosive Hazard Detection in FLGPR, and the fusion of FLGPR and EO imaging sensors. We built and distributed software to register FLGPR and imagery, register imagery to ground truth UTM coordinates, and collect class-based object features from image sequences. Our approaches are described below.

Forward Looking Anomaly Detection via Fusion of Infrared and Color Imagery

We investigate two algorithms for the detection of interesting and abnormal objects in color and infrared imagery taken from a moving vehicle observing a fixed scene. The purpose of detection is to cue a human-in-the-loop detection system, thereby alerting an operator to areas that require human inspection. This vehicle based detection system is used for clearing hazards from roads. It incorporates two wide field of view (WFOV) cameras, one color and one un-cooled (non-polarized) long wave infrared (LWIR), which is simply referred to as IR here, as well as three zoomable, narrow field of view (NFOV) cameras which the vehicle operator is able to use for closer inspection of specific locations. The first algorithm is based on change detection, utilizing previously captured color and IR imagery of the lane that is known to be free of hazards. The second algorithm is based on direct detection and does not utilize any prior information about the lane other than the incoming imagery. The data used here comes from a data collection at a US Army test site. Change detection focuses specifically on a single lane which will be referred to as lane B. It contained sixteen total hazards, three of which were buried.

Captures traveling both east and west on the lane were made. Direct detection also looks at color imagery of lane B captured from a second vehicle, referred to as system B, with the same model color camera used in the detection system previously described, referred to as system A.

CHANGE DETECTION

For change detection, imagery of lane B captured on separate days was used. Targets of interest were present on one of the days and not present on the other. The time of day of the two data captures differed by approximately thirty minutes, mid-afternoon, and the weather conditions were similar. As mentioned in the introduction the imagery comes from two WFOV cameras, one un-cooled long-wave IR and one color, mounted in fixed locations on a moving vehicle. The color imagery was uncompressed 32-bit, ARGB with a resolution of 1024x768 and the IR imagery was 8-bit, grey scale, with a resolution of 640x480. The cameras were synced, and captured at a rate of fifteen frames per second. The vehicle traveled at approximately fifteen miles per hour. Example images are shown in figure 1.

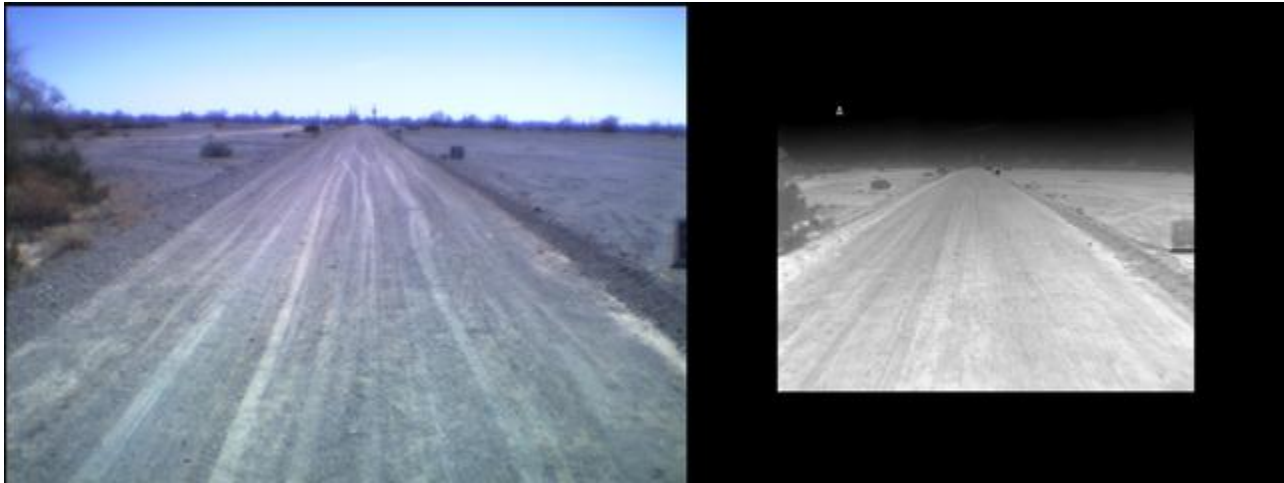


Figure 1. Left: WFOV color image. Right: WFOV long-wave IR image.

Selecting background images

Given a frame from the current run, the first step in the change detection algorithm is to select a set of background frames with which to compare. Originally, this was intended to be done using image space to world coordinate mappings, as described in last year's report, allowing fast computation of the overlap between the current frame and each frame from the background run. This would allow efficient frame selection, computation wise, and ensure that the background frames with the best view of the detection area within the current frame were chosen. However, the captured data lacked the necessary heading and GPS information to allow these computations. Since this information will be available in future systems, a simple alternative involving manual intervention was used for tests here. Namely, the closest matching frame in the background run, in terms of overlap amount, for every hundredth frame in the current run, there were approximately two thousand frames per run, was manually selected and recorded. Linear interpolation was then used to compute the closest matching background frame for the rest of the frames in the current run.

Another question that arises is how many background frames should be selected for comparison? Again, given the image to world coordinate mappings it is simple to determine how many background frames overlap the current frame and the differences in their respective viewing distance and angle. This information could be used to intelligently determine the number of background images to use. Since this was not possible, an arbitrary fixed number, five frames, was used for tests here. The five frames consisted of the nearest matching frame, n , determined as explained in the previous paragraph, plus frames $n-2$, $n-4$, $n+2$, and $n+4$ from the background run.

Image to image transformation

Once the set of background frames has been selected each is mapped into the current frame image space using affine scale-invariant feature transform (ASIFT) keypoint correspondences and perspective transformation. ASIFT keypoints were

chosen due to their invariance to scale, rotation, translation, and viewing angle. The invariance to viewing angle separates ASIFT from Lowe's popular scale-invariant feature transform (SIFT) keypoints. This difference is important because it is likely that the viewing angle difference between some background frames and the current frame is large enough that SIFT keypoints are inadequate. This was the case for the data used here. Given the keypoint pixel location correspondences the parameters of the perspective transformation [A-H], as defined in equation 1 below, which maps a pixel location in a background image (X_b, Y_b) to a pixel location the current image (X_c, Y_c), were computed using least trimmed squares (LTS) regression with seventy-five percent of the correspondences considered to be good. LTS was chosen for its robustness in the presence of correspondence mismatches as compared to traditional least squares regression.

$$\begin{bmatrix} X_c \\ Y_c \\ 1 \end{bmatrix} = \begin{bmatrix} A & B & C \\ D & E & F \\ G & H & 1.0 \end{bmatrix} \begin{bmatrix} X_b \\ Y_b \\ 1 \end{bmatrix} \quad \equiv \quad X_c = \frac{A * X_b + B * Y_b + C}{G * X_b + H * Y_b + 1}, \quad Y_c = \frac{D * X_b + E * Y_b + F}{G * X_b + H * Y_b + 1} \quad (1)$$

Lighting and contrast adjustment

Even though the background runs were captured at roughly the same time of day and under similar weather conditions as the current runs there was considerable lighting and contrast difference in both the color and IR imagery. While these differences were relatively static over the length of a run for the color imagery, they changed on a frame by frame basis for the IR imagery. This is not surprising given the nature of un-cooled long-wave IR. To compensate for these differences gain and offset adjustments for each channel (R, G, and B for color and Y for grey scale IR) were computed.

For color, since the adjustment was constant over the length of a run, only the first few frames of the sequence were used to compute the gain and offset for the R, G, and B channels. The parameters were estimated using separable CMA-ES such that the Euclidean distance between the 3D color histograms of the images from the current run and the adjusted background images was minimized⁵. For the 3D color histogram a quantization step size of eight was used, resulting in 32^3 total bins. This was done before computing the ASIFT keypoint correspondences. For IR, gain and offset were computed on a per frame basis via linear regression of the pixel values in the background image on to the pixel values of the current image. The linear regression was done after performing the image space mapping, and only performed within the desired detection window of the current frame.

Difference image

Once the background frames are mapped into the current image space and have been adjusted for lighting and contrast, each is differenced on a per pixel basis with the current frame within the detection window. This differencing is performed using Euclidean distance in CIELAB color space for the color imagery and Euclidean distance between grey levels for the IR imagery. Use of CIELAB color space was motivated by its superior perceptual uniformity compared to RGB and slight illuminant invariance, as explained in our previous work. Use of CIELAB color space involves a number of parameters that are usually neglected, specifically the gamma function used to convert from linear to gamma-corrected RGB, the CIEXYZ tristimulus values of the red, green, and blue primaries and the white point defining the RGB images' color gamut, and the illuminant under which the image was originally captured. Unfortunately, the first two items were not known for the data used here, but intelligent guesses were made based on image resolution. Specifically, sRGB primaries, white point, and transfer function were used. D65 (noon-daylight) illuminant was used, and corresponds closely to the capture conditions.

Once the individual difference images have been created the combined result is obtained by taking the minimum difference at each pixel location. This image is then low-pass filtered using a simple averaging window of size $N \times N$. The effect of different values of N is investigated later. After low-pass filtering the image is thresholded at a value 'T', and a morphological flood fill operation, starting from the edges and filling background pixels inwards, is performed to close any holes. The resulting binary mask is then used for target declaration.

Target declaration

Given a binary mask image target declarations are made by finding all connected components in the image using four-neighbor connectivity. These connected components are then added to a linked list which is sorted by size, i.e. the number of pixels in each connected component. All connected components with size less than 'C' are removed. This eliminates detections too small to be actual targets. Next, all connected components whose centroid is within 'D' pixels of a

larger connected component's centroid are merged with the larger connected component. The centroids of the remaining connected components are declared as target locations.

One drawback of this method is that large blocks, i.e. detected areas in the binary image, will yield only a single target declaration. When a connected component covers a significant horizontal or vertical span, especially in an irregular shape, the centroid is generally not a good location for target declaration. This is more of a problem in the direct detection method described later than for the change detection algorithm. To handle these scenarios, connected components are not allowed to span more than $2 \times D$ pixels. Such connected components are arbitrarily split into two or more, smaller connected components based on the order in which pixels were added. An example of target declaration from a binary mask is shown in Figure 2 below.



Figure 2. Target declaration from binary mask. Red dots are target declarations.

Scoring

Due to the lack of heading and GPS information scoring based on GPS ground truth location was impossible. Therefore, manual image truthed scoring was used. Image truthing consisted of choosing a subset of frames from each run and having a person manually label each of those images by selecting the center-point, in pixel coordinates, of any targets present. The subset of frames was selected by finding the last frame in which each target could be seen and then selecting every previous fourth to sixth frame until the target was far enough away that a human could no longer distinguish it as a target. For buried targets, since they could not be seen visually, the selected center-point was based on fiducials placed nearby.

Given this image truth information, a lane was scored by taking the target declarations for each of the image truthed frames and computing the number of false alarms and correct detections. A false alarm was any target declaration not within the halo distance 'H', in pixels, of a target center point location. All target declarations within the halo distance 'H' of a target center point location were counted as correct detections. A target was said to be detected if at least one correct detection corresponded to it. Since the same world location was typically seen in multiple image truthed frames, and no linking of false alarms was performed, the number of false alarms was typically higher than the number of physical false alarm locations using this method. However, it provided a rough means to objectively test the algorithms and the effect of different parameters.

Lane B contained sixteen total targets, three of which were buried. For the east run, referred to as B East, the image truth consisted of 116 frames with no more than one target per frame. Eighteen of those frames contained buried objects. For above ground only scoring those frames were ignored, resulting in 98 image truthed frames. For the west run, referred to as B West, the image truth consisted of 111 frames with no more than one target per frame. Sixteen of those frames contained buried objects, resulting in 95 image truthed frames for above ground only scoring. For IR scoring, the IR images were transformed into color image space and scored using the same image truth used to score the color images. This transformation was performed using perspective projection, as given by equation 1. Since the location of the two cameras on the vehicle never changed this transformation was fixed.

Color change detection results

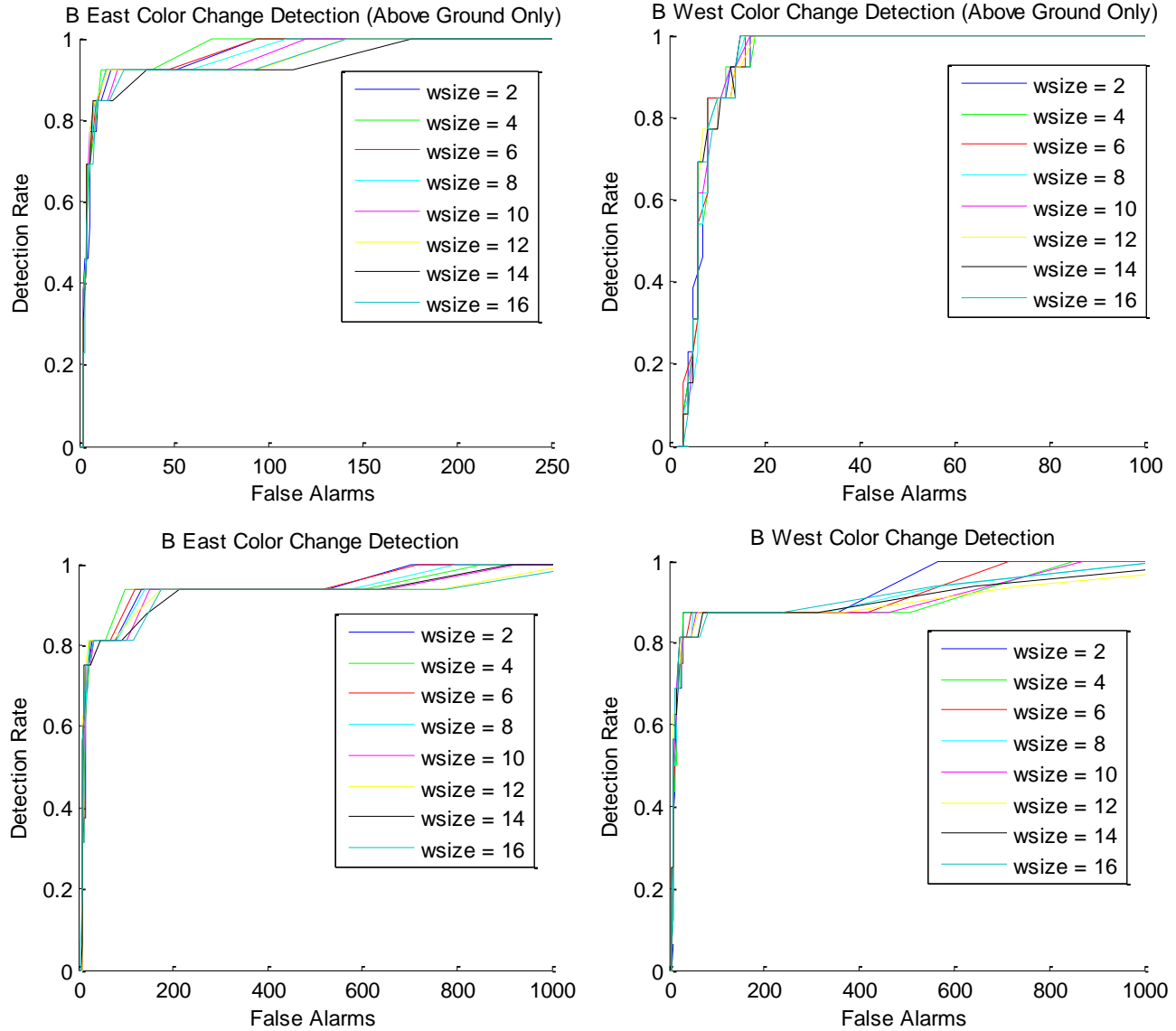


Figure 3. Color change detection ROC curves.

Color change detection results are shown in Fig. 3 for lane B for both east and west directions. The top two curves show results for above ground targets only, while the bottom two include buried targets. The graphs are pseudo-ROC curves showing detection rate on the y-axis and number of false alarms on the x-axis, instead of false alarms per some unit measure, as the detection threshold ‘T’, from section 2.3, was varied. The different lines on each plot represent different averaging window sizes ranging from two to sixteen. Values of $C=25$ and $D=50$, as described in section 2.4, were used for scoring, as well as a halo size of 50 pixels. The detection window was between scan lines 200 and 540.

For above ground targets the behavior on B East and B West was similar except for detection rates above ninety percent where the number of false alarms was substantially higher on B East. The curves suggest that most of the targets were easy to detect, but one or two did not create a significant difference when compared to the background images. The averaging window size had little effect at most detection rates. As expected, significantly more false alarms must be accepted to detect the buried targets as indicated by the bottom two plots. In general we do not expect color to detect buried targets unless little or no attempt has been made to hide the signs of digging and disturbed earth.

IR change detection results

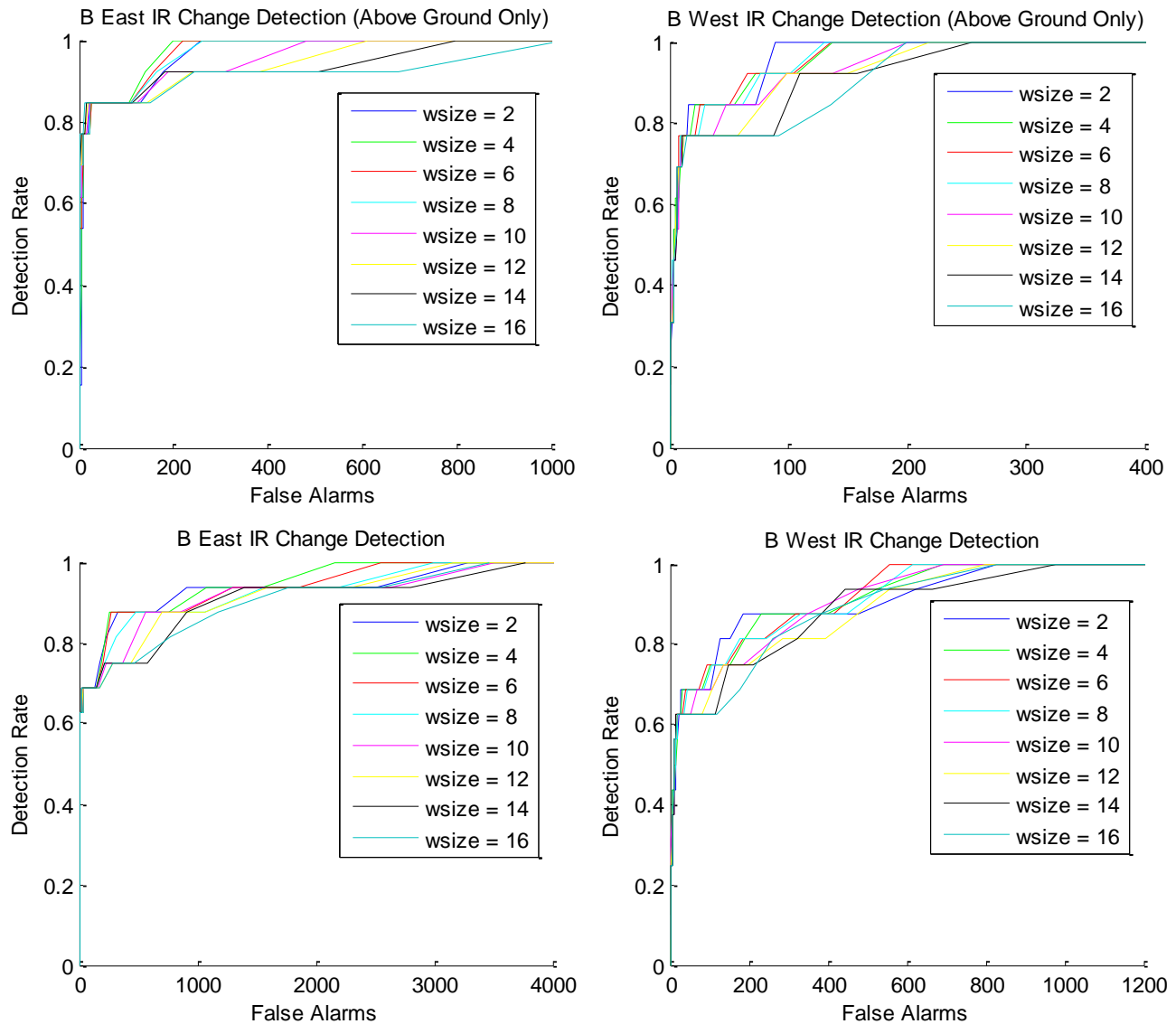


Figure 4. IR change detection ROC curves.

IR change detection results are shown in Fig. 4 for lane B for both east and west directions. The top two curves show above ground only results, while the bottom two curves include buried targets. The same values of C , D , and halo size as used for the previous color change detection results were used. The detection window was between scan lines 120 and 330.

As with color, the averaging window size had no distinct effect at detection rates of roughly eighty percent or lower. However, at higher detection rates window sizes in the range of four to eight performed best. Clearly IR has more false alarms for detection rates above seventy-five to eighty percent than color. This is especially true in the buried target case. While these results indicate that alone IR does not perform as well as color, it is possible that the two could be combined to give better results than either individually. In fact, inspection of the binary masks makes it clear that the color and IR change detection pick up different effects. It is hoped that actual targets will cause changes in both color and IR. This is likely for above ground targets since they produce visible changes detectable in color, and most of the targets have emissivity and absorption values different than the local surroundings making them detectable in IR.

Color and IR change detection fusion

Next, the fusion of color and IR change detection was investigated. Applying the logic from the preceding paragraph, that we expect true objects of interest to cause changes in both color and IR, a simple AND operator fusion was used. The binary masks output for color and IR change detection were combined by taking the minimum at each pixel location. This new mask was then passed to the target declaration routine and scored using image truth in the same manner as the previously presented results.

Since this algorithm contains two thresholds, ‘T’ for color and ‘T’ for IR, displaying a single ROC curve is not possible. Instead, the two thresholds were varied and for each detection level the best result in terms of number of false alarms was selected. These points are shown in the plots in Fig. 5. The same averaging window size was used for both color and IR when fusing. Multiple window sizes were tested for fusion. Specifically, window sizes of two, four, six, and eight. For comparison, each plot also contains the curves for color and IR change detection alone with a window size of four.

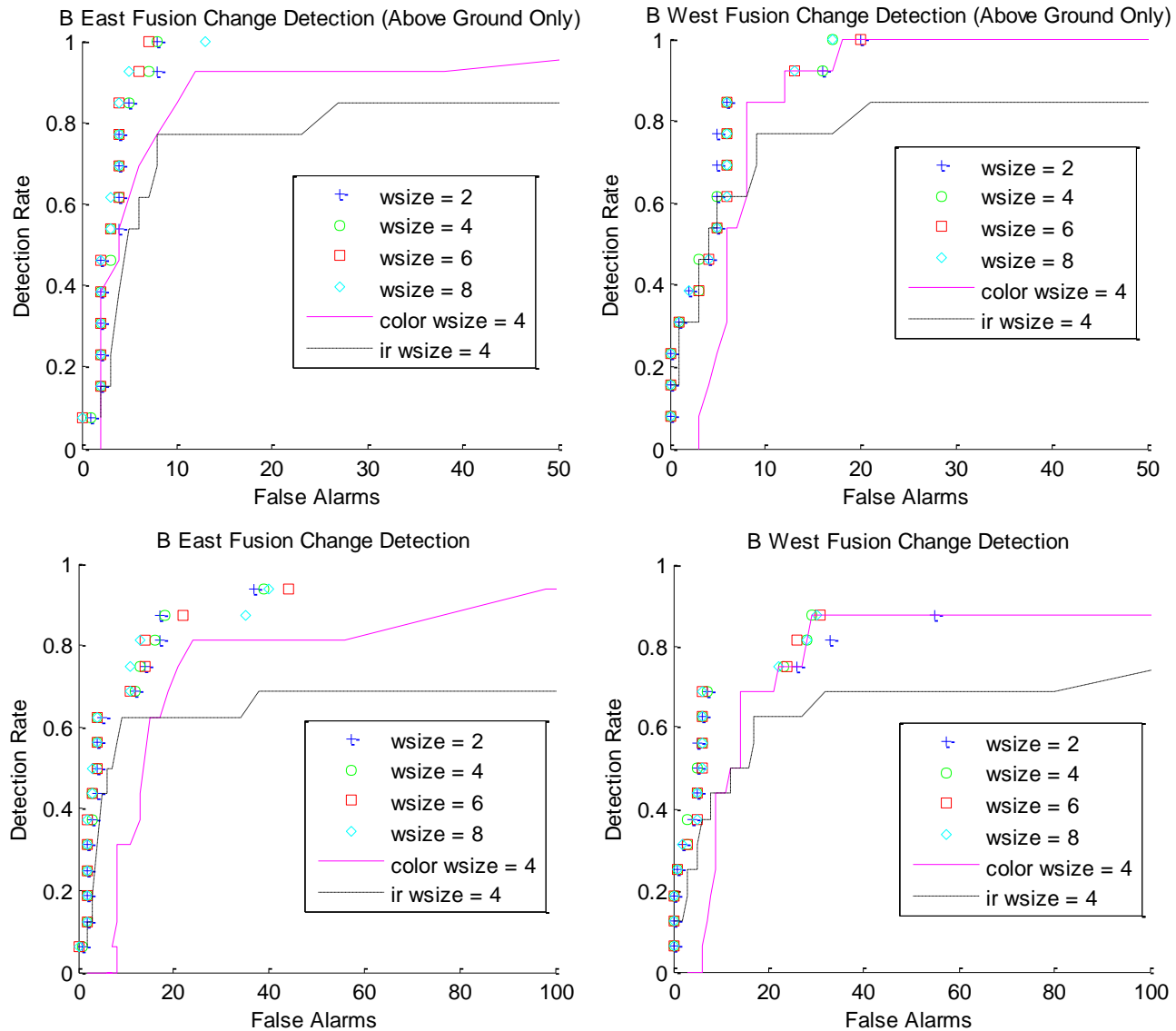


Figure 5. Change detection color and IR fusion.

For above ground only detection fusion is able to reduce the number of false alarms substantially on B East for high detection levels. Fusion doesn't help much on B West. However, color only change detection alone already produced few

false alarms. For a one hundred percent above ground only detection rate the best color change detection can do is 70 false alarms on B East and 15 false alarms on B West. The best IR change detection can do is 200 false alarms on B East and 89 false alarms on B West. Fusion is able to achieve 7 false alarms on B East and 17 false alarms on B West.

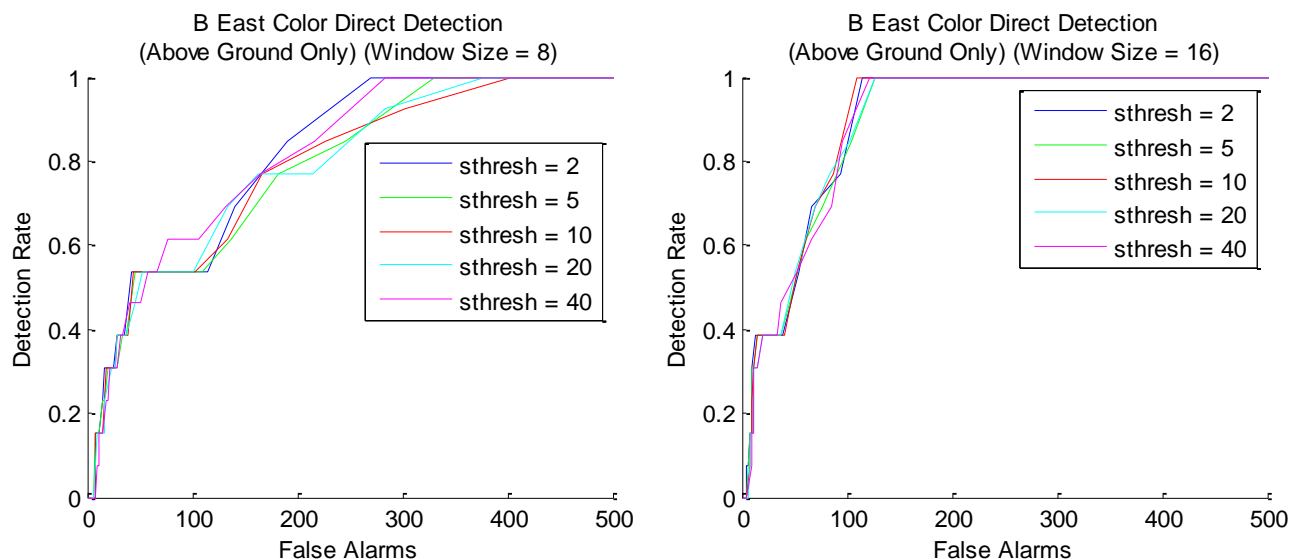
DIRECT DETECTION

Direct detection and testing utilized the same data as change detection, simply not making use of the background runs. Again, the idea is to cue a human-in-the-loop detection system, thus direct detection is attempting to detect interesting or unique parts of the image. It does not attempt to look for specific types of objects since the forward looking anomaly task is not well defined and the types of targets vary. Essentially, the direct detection algorithm described here is intended as a pre-screener which would be coupled with other detection and/or classification algorithms.

Image self-similarity

Direct detection uses the concept of image self-similarity. First, a detection window within the current frame is selected based on the desired detection range. This window is then broken into overlapping blocks of size $N \times N$. Each of these blocks is exhaustively compared to every other $N \times N$ block in the image, not limited to blocks within the detection window. Blocks within a small region around the current block are not used for comparison. The idea is that any interesting objects should be unique, i.e. a block containing such an object will look different than any other block in the image. Whereas blocks that don't contain interesting objects, i.e. background such as ground, bushes, etc..., will look similar to other blocks in the image. Block comparisons are performed using mean Euclidean distance in CIELAB color space for color imagery, and mean Euclidean distance between grey levels for IR imagery. Again, the choice of CIELAB color space for color is motivated by its superior perceptual uniformity compared to RGB. If a block has a distance greater than 'T' to another block they are said not to match. If a block under consideration in the detection window does not match at least 'S' other blocks in the image then it is flagged as interesting and the corresponding area in the image is marked in a binary detection mask. The full detection mask for the frame is the combination of all flagged blocks. The exhaustive search process is computationally demanding even with efficient implementation as described in [6]. This could probably be replaced with a faster search strategy such as diamond, square, or hexagonal search initiated at many evenly spaced points in the image. However, we use it here for initial algorithm testing.

Direct detection results



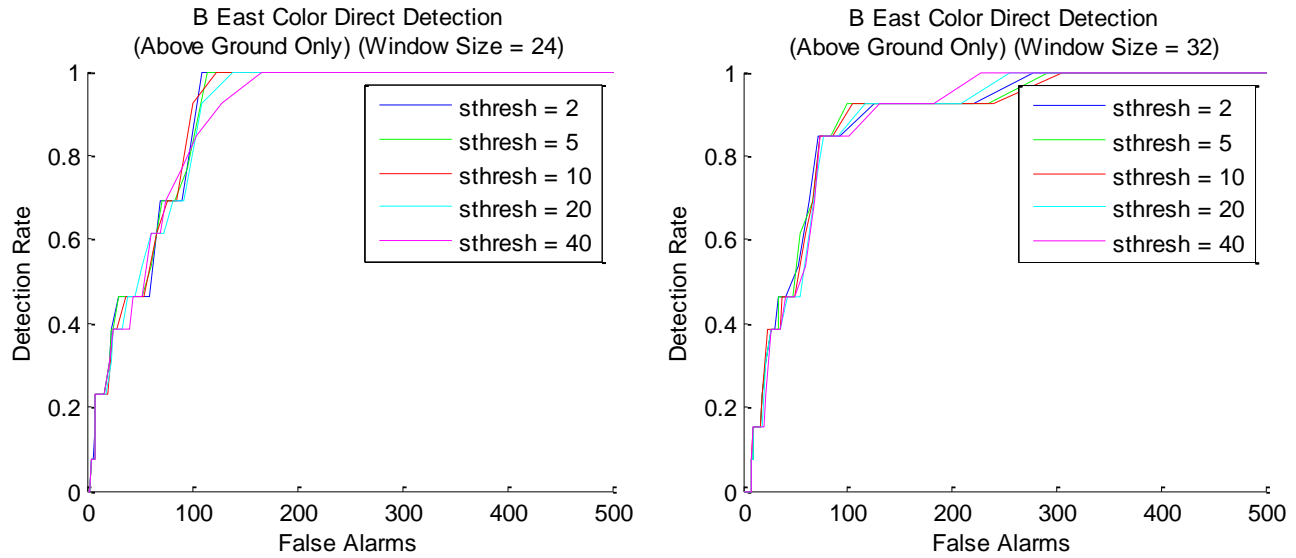
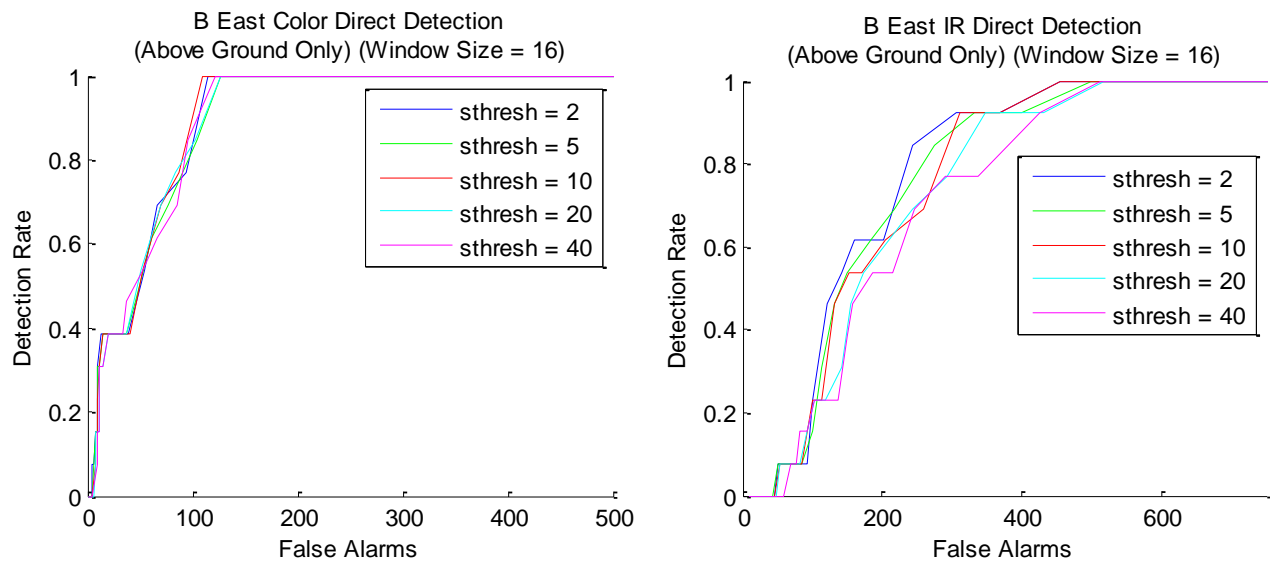


Figure 6. Color direct detection results for B East.

Fig. 6 shows above ground target direct detection results for color on B East. Each plot corresponds to a different block size, or window size, N . Within each plot, each curve corresponds to a different 'S' value as the distance threshold 'T' was varied. The halo size, C , and D values used for scoring were the same as used for change detection. The detection window was between scan lines 304 and 580 for color and between 184 and 364 for IR. These plots indicate that the 'S' value has little effect. This is not surprising since it is tightly coupled with the difference threshold 'T'. Much of the effect of increasing 'S' can be produced by decreasing 'T'. Block sizes of sixteen and twenty-four perform better than either eight or thirty-two. These trends are also seen with the IR imagery (plots emitted for space).

Results for color and IR on B East and B West using a block size of sixteen are shown in Fig. 7. Again, the 'S' value has little effect. Obviously, direct detection has more false alarms than change detection, but for such a simple algorithm intended as a pre-screener the results are encouraging. Color performance is significantly better than IR indicating that most of the targets stand out, in terms of image self-similarity, much better in color than IR. The lower resolution and sharpness of the IR imagery plays a part in this. Further investigation is needed to determine whether IR would perform better at different times of the day, such as sunrise, or whether a single (non-polarized) LWIR band is unable to detect differences between the types of targets and background environment present in this data.



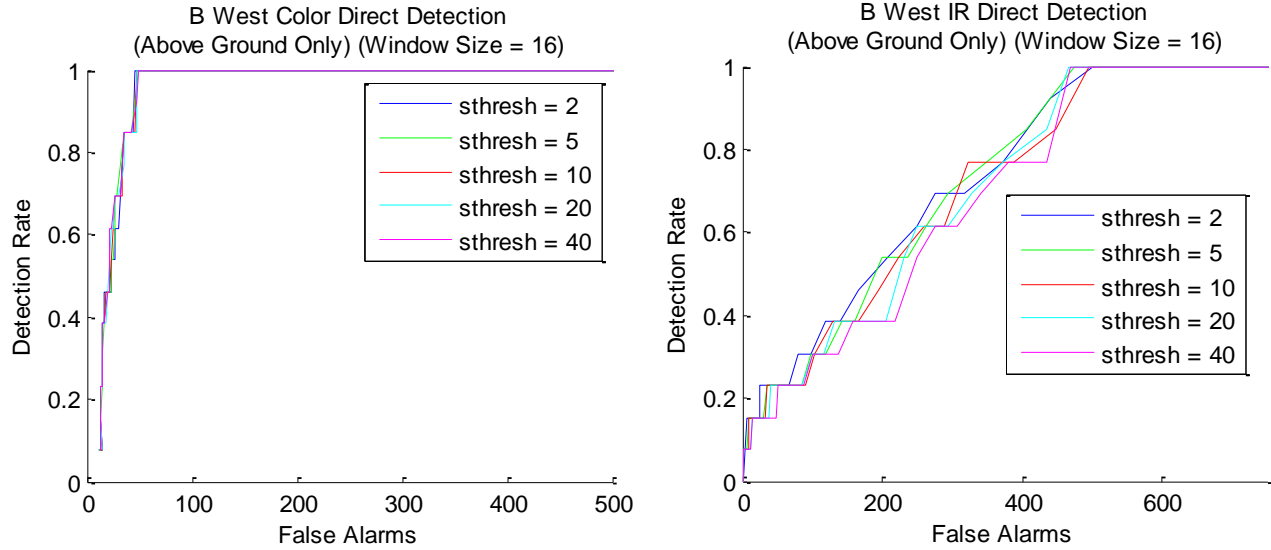


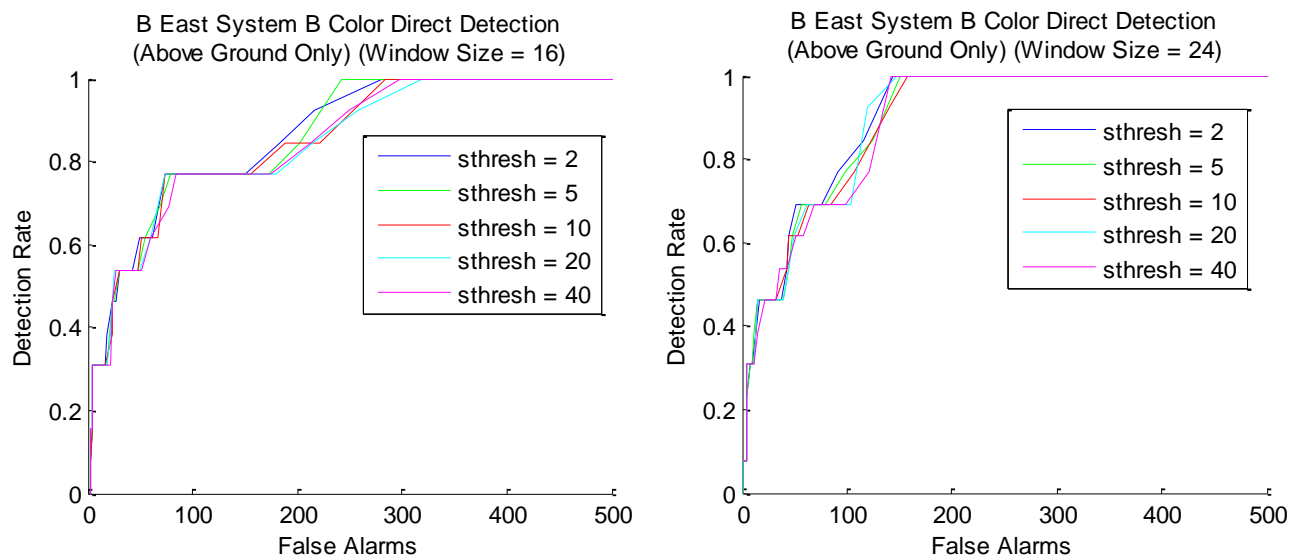
Figure 7. B East and B West direct detection results for color and IR with block size sixteen.

Fig. 8 shows lane B direct detection results for the color imagery from detection system B. Only above ground targets were considered. Results for block sizes of sixteen and twenty-four are shown. For system B, the image truth for lane B East consisted of 87 frames and the image truth for lane B West consisted of 91 frames. All scoring parameters were the same as those used for the system A results. The one difference between color direct detection with system A versus system B was that the detection window was between scan lines 200 to 600 for system B due to the steeper downward angle of the camera.

As with system A, performance is better on B West than B East. Detection rates above eighty percent on B East show substantially fewer false alarms for block size twenty-four than sixteen. This is not seen on B West. Overall, the results from system B are similar to the color direct detection results of system A.

Direct detection color and IR fusion

As with change detection the fusion of color and IR direct detection was investigated. The fused binary detection mask was obtained by taking the minimum between the color and IR detection masks at each pixel location.



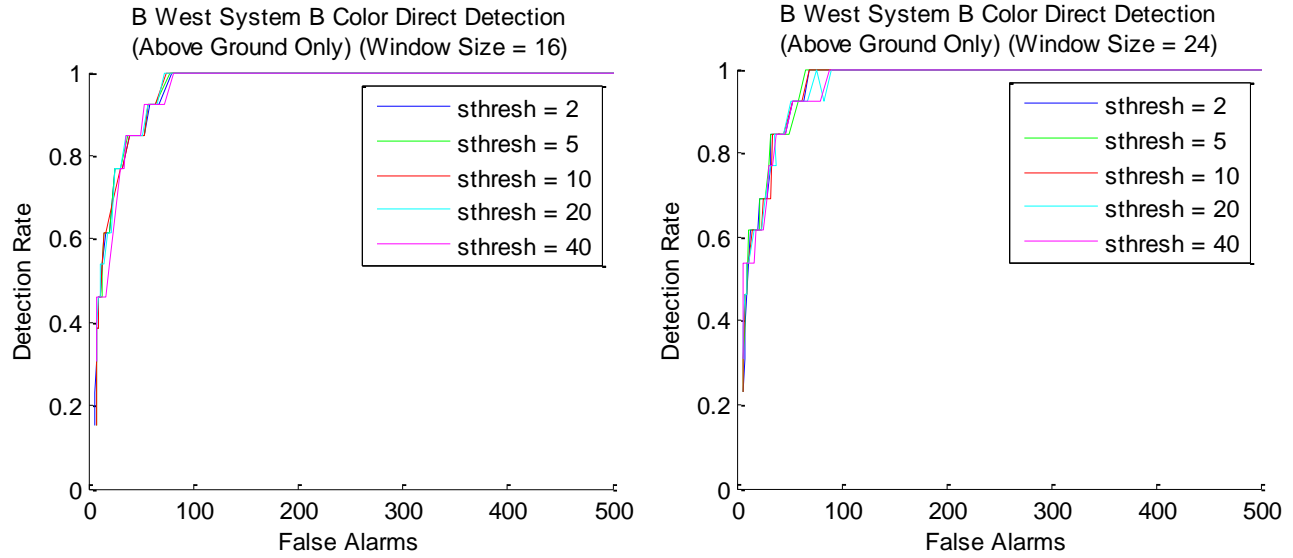


Figure 8. B East and B West direct detection results for system B.

Since the 'S' parameter value was shown to have little effect $S=10$ was used for all experiments. When fusing, the same block size was used for both color and IR. The results for block sizes of sixteen and twenty-four are shown for B East and B West in Fig. 9. Since two thresholds are being varied, 'T' for color and 'T' for IR, the best point for each detection rate, based on number of false alarms, is shown. For comparison, the curve for color alone with block sizes of sixteen and twenty-four is also shown on the graph. Fusion reduces the number of false alarms on both B East and B West.

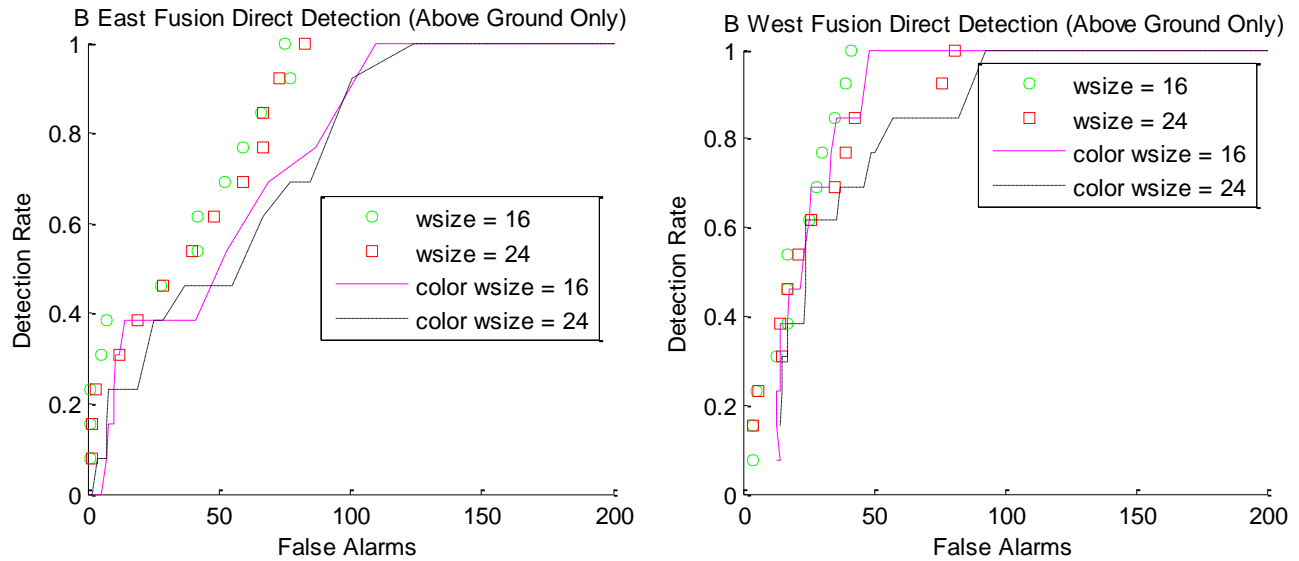


Figure 9. Direct detection color and IR fusion results for B East and B West.

FULL LANE TESTS

Finally, change detection and fusion of change and direct detection were evaluated using B East as a training lane to select thresholds and B West as a testing lane. Based on the results from section 2.9 for B East, a window size of four and color and IR thresholds of thirty-two and twenty-two were selected. These thresholds gave one hundred percent above ground target detection for a window size of four on the B East image truth. Using these settings on the entire B East run, 1910 frames spanning roughly 0.9km down track, resulted in 456 total detections. All thirteen above ground targets were detected, plus one of the buried targets. 248 of the detections were on actual targets. 49 were on three manmade objects that were not

present, or were positioned differently, in the background run, leaving 159 unlinked false alarms. It is likely that many of these false alarms correspond to the same physical location on the ground. However, this was not verified as it would have required significant manual labor due to the lack of GPS information. Using the same settings on the B West run, which contained 1742 frames, resulted in 445 total detections. Eleven of the thirteen above ground targets were detected. None of the buried targets were detected. 203 of the detections were on actual targets. 63 of the detections were on the three manmade objects mentioned previously, leaving 179 unlinked false alarms.

For fusion of direct detection and change detection, change and direct detection within each modality, color or IR, were first combined by taking the minimum of the associated binary masks at each pixel location. A flood fill was then performed on the result using the original change detection mask as the seed. This step restored any connected component in the change detection mask that had at least one pixel survive the fusion with direct detection. After this process was performed for each modality, the two resulting binary masks were combined by taking the minimum at each pixel location. The same settings were used for change detection as were used in the previous paragraph. For direct detection a block size of sixteen, 'S' value of ten, and color and IR thresholds of thirty and fourteen were chosen based on results for B East from section 3.3. On B East this resulted in 175 total detections. Twelve of the thirteen above ground targets were detected. None of the buried targets were detected. 97 of the detections corresponded to actual targets, 33 corresponded to the three manmade objects, leaving 45 unlinked false alarms. Using these settings on B West resulted in 258 total detections. Ten of the thirteen above ground targets were detected. None of the buried targets were detected. 140 of the detections corresponded to actual targets, 50 corresponded to the three manmade objects, leaving 68 unlinked false alarms.

Anomaly Detection in Forward Looking Infrared Imaging Using One Class Classifiers

There are several common challenges that any anomaly detection algorithm is faced with:

1. the number of abnormal objects (road hazards) is several order of magnitude smaller than the number of background objects. This problem is sometimes called "the class imbalance problem";
2. the characteristics of "future" abnormal objects might be very different from those available in the training set (previously seen);
3. the background objects may change over time. At first they will probably appear as "anomalies". Although some authors differentiate between "anomaly detection" and "novelty detection", we believe that the resulting algorithms are, in essence, similar.
4. when the abnormal objects are the results of malicious actions, they are often made to appear as part of the background (camouflage);
5. although noise is sometimes treated as "anomaly", it is a non-interesting anomaly. Moreover, its presence complicates the task of finding the interesting ones;

Extra challenges encountered in processing IR images from vehicle mounted camera:

6. the image perspective: both normal and abnormal objects look different depending on the distance from the vehicle.
7. the physics of IR imaging: both normal and abnormal objects look different depending on the time of the day and the outside temperature.

To address challenges 1 and 2 above, we used an anomaly detection approach called one class classifier (OCC) to learn the background objects (e.g. road, bushes, rocks, etc.). OCCs are a type of classifiers that do not require two classes for training. Here, we used the following ones: one class (spherical) support vector machine (OCSVM), one class nearest neighbor (OCNN) and one class Gaussian mixture (OCGM). We plan to address challenge no. 3 by adaptively training of OCCs, but we did not provide any results here. Challenge no. 4 is addressed by fusing the results obtained by the IR sensor with other imaging modalities such as color imagery and forward looking GPR. One way of addressing challenge no. 5 is to use temporal fusion: objects not identified in at least m of n consecutive images, $m < n$, are discarded. Challenge no. 6 was

addressed by using the perspective transform in object tracking. Here we used the method described in to account for perspective..

One important problem for our classifier is choosing the IR image properties to use for anomaly recognition. We note the distinction we make here between IR image properties that are related to physical phenomena and IR image features that are a mathematical representation (i.e. image processing) of the reality. Possible IR imagery properties for detecting buried road objects are: surface texture, spectral signature of the disturbed earth and differences in thermal inertia. Local texture variations of the surface above a buried road hazard can be used for detection only if the image is taken soon after the object is placed in the ground. After a while, weathering or animal traffic may change the surface texture. Here we do not use soil texture features due to vehicle traffic over the buried objects area (road). For the same reason, features based on the spectral signature of the disturbed earth are of limited use in our case. The thermal inertia is useful in parts of the day when there is rapid change in temperature such as dawn and dusk. Since our experiments were performed mid day, thermal inertia based features are not useful. Moreover, due to the experimental conditions (road traffic, mid day, weathering) only above ground metal objects can be detected. During a sunny day, in the absence of temperature variation, a wide array of objects such as bushes, pieces of metal, cacti, rocks, etc., appear bright in IR imagery. Consequently, for these test conditions, the only available physical property for differentiating between background and road hazards was the shape of the objects.

Among the many algorithms used for finding abnormal regions in IR imagery such as matched filter, clustering-based, mathematical morphology-based, the ones based on the RX algorithm seem to be the most popular. The RX algorithm, mainly a multispectral method, is based on computing a confidence value that the center pixel from an image set window comes from the distribution of the pixels in that window (i.e. background distribution). If the confidence is low, we may deal with an anomaly. In detection applications using airborne IR imaging, the objects in the field of view maintain a relative constant size, hence the size of the sliding window employed for computing the local distribution may be constant. By contrast, in our application, the image perspective causes objects farther from the vehicle to appear smaller than the ones that are closer. While, technically, the window can be adjusted to match the perspective for a given target size, its area may become too small for computing a meaningful feature distribution. Moreover, the availability of images in only one IR band (8-12 μm) made the RX algorithm less suitable for our application.

The goal of our algorithm is to cue the operator of a vehicle of abnormal objects present in the environment. The proposed cuing method has two steps. First, for each IR frame we generate a set of possible points of interest using a corner detection algorithm. Then, we employ an OCC to remove the hits associated to "normal" objects such as bushes, road, road side and shadows.

CORNER DETECTION ALGORITHM FOR SHAPE CAPTURE

Corners often contain critical information about the objects they belong that can be used in the identification task. The corner detector used in this work is a multiscale algorithm based on curvature scale space (CSS) calculation. The main steps of the algorithm are as follow:

1. Apply Canny edge detection to each IR image
2. Extract edge contours from the edge map. Optionally, small gaps in the contours can be filled. Mark the end points of the open curves. The algorithm can identify corners on both sets of curves. However, in this work, we chose to use only close contour curves.
3. For each contour identified above compute the curvature at a given low scale. This approach will generate many corner candidates (local maxima of the curvature).
4. False corners are eliminated based on global features such as average curvature, corner angle and ratio of the axes of the inscribed ellipse, computed over the region of support (the contour bounded by the two nearest curvature minima). While the average curvature is computed dynamically for each region of support, the maximum corner angle and the maximum ellipse axes ratio are user inputs.

An example of the algorithm output is shown in Figure 10.

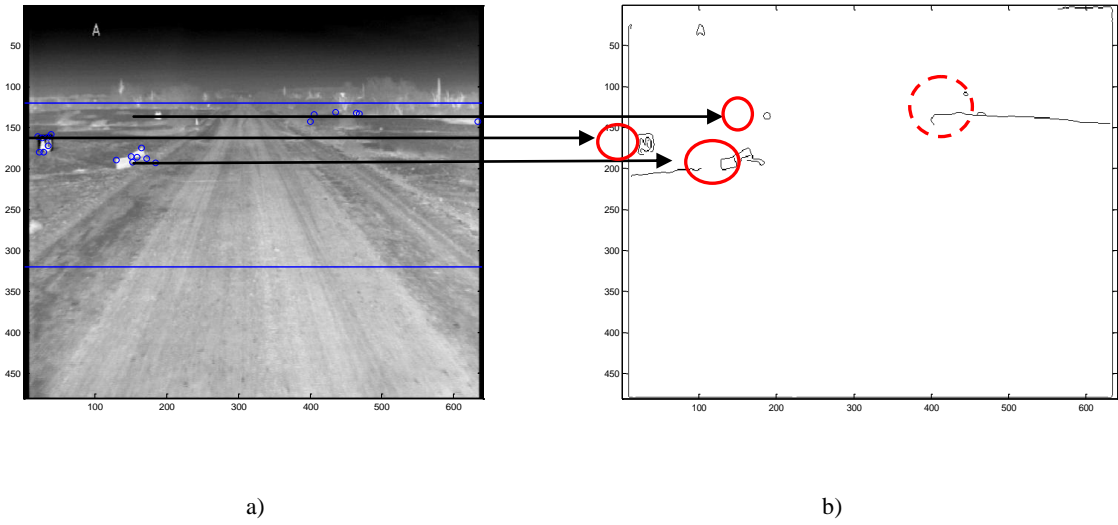


Figure 10. Example of the corner detection algorithm for an IR image (a). The corner detector uses the edge map generated by the Canny algorithm (b).

Figure 10.a shows a typical 640×480 IR image from our dataset. Our detection algorithm is restricted to run in a horizontal band (blue lines) between $y=120$ and $y=320$. Figure 10.b shows the output of the Canny edge detection for the image from Figure 10.a where the objects of interest are circled: two metal fiducials (one around (30, 170) and another around (200, 130)) and a group of man-made objects (located around (200, 140)). Some other closed contours are formed by bushes and shadows (far right side of the road, circled with dashed line in Figure 10.b).

There are two key steps in the corner detection algorithm: edge generation (step 1) and false corner elimination (step 4). Step 1 is controlled by the high (H) and the low (L) threshold of the Canny edge detector. Higher values of H and lower values of L produce more edges. While, ideally, the $\{H, L\}$ values should be different for each frame, here we kept them constant for a given run, i.e. $L=0$ and $H \in [0.15, 0.35]$. Step 4 has five main parameters:

- C, the ratio of the axes for the corner inscribed ellipse. We used $C=1.5$ and $C=1$ (rounder corners). For example, the corner detection algorithm in Figure 10 was run with $C=1.5$. For this reason, the fiducial at (200, 140) is not detected (it looks like a circle, hence $C \sim 1$).

- T, maximum angle of a corner. A higher T value would produce more hits for each object. We experimented with values around $T=160$;

- Sigma, a contour smoothing parameter. We used $\text{Sigma}=3$;

- Endpoint, whether to consider or not edge endpoints as corners. We used $\text{Endpoint}=0$ (we only wanted closed contours).

- Gap_size, the number of pixels required to close an open contour. Given the average size of our road hazards (0.5 m) we chose $\text{Gap_size}=20$ which represents about 1 m in the middle of our processing window. Ideally, the gap size has to account for perspective in an image (see Figure 11).

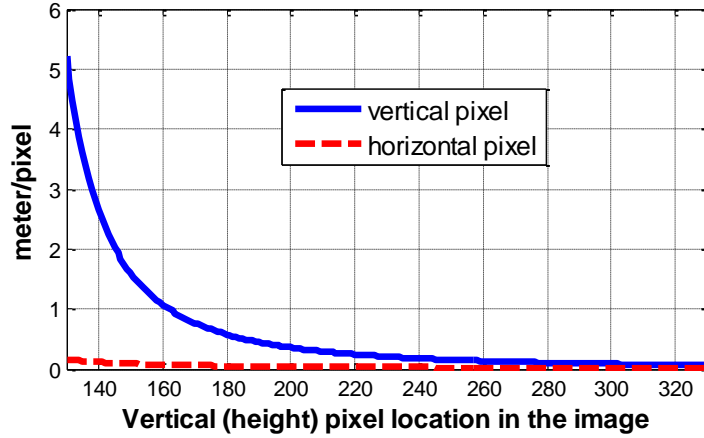


Figure 11. Correspondence (meter/pixel) between the real word dimensions and horizontal (dashed) and vertical pixels (continuous) along image height. For example, at $y=160$ a vertical pixel represent about 1 m whereas an horizontal one is about 0.1 m wide. Furthermore, an 0.5 m object would look about 3 pixels wide at $y=130$ and about 30 pixels wide at $y=320$.

Three of the 5 false alarms from Figure 10.b (dashed circle) come from closing far away contours (bumps in the horizontal bush line) with too large of a gap size (at $y \sim 130$, 20 horizontal pixels represent around 4 m). Here, however, we kept Gap_size constant along the field of view.

A MATLAB implementation of the above algorithm, *corner.m*, can be downloaded from: <http://www.mathworks.com/matlabcentral/fileexchange/7652-a-corner-detector-based-on-global-and-local-curvature-properties>.

As we will describe next, we will further reduce the number of false corners by employing an one class classifier (OCC). However, even after this reduction, it is possible that multiple corners per object are detected. This characteristic of the current algorithm is alleviated by temporal fusion. However, it still can lead to a disproportionately large number of false alarms, i.e. greater than the number of detected objects. We are currently working at a local clustering algorithm that would replace the corner hits of an object, with a single hit located in the center of the cluster (see for example the fiducial located at (30, 170) from Figure 10.a that has 7 corner hits).

ONE CLASS CLASSIFIERS

Here we examine three OCCs: OC Gaussian mixture (OCGM), OC nearest neighbor (OCNN), and OC support vector machine (OCSVM). The OCCs can be classified in boundary methods (such as Parzen and Gaussian mixture) and boundary methods (such as nearest neighbor and SVM). The unified approach to OCC does not use a threshold for accepting the normal class objects. Instead, it assumes that a certain percentage, t_0 , of the training data are outliers. This approach is also based on the additional assumption that the outliers (the second class, i.e. surface road hazard) is uniformly distributed in feature space "around" the "normal" class (i.e. background). The main effect of this formulation is that it eliminates from the training set some unusual "normal" class objects, that might exist in the training set, for example, due to noise. Also, it make it easier to define a unified threshold for all classifiers used in an application.

The OC Gaussian mixture (GM) method, OCGM, is mathematically similar to the traditional one, that is:

$$OCGM(\mathbf{x}) = \sum_{i=1}^M w_i N_i(\mathbf{x})$$

where \mathbf{x} is the feature vector extracted from corner s , w_i are a set of weights and N_i a set of Gaussians functions. Here we use six features: average and standard deviation of the gray level, horizontal and vertical gradient, respectively. The features are calculated in a 3×3 neighborhood around each detected corner. As mentioned above, instead of using a (probability) threshold p_0 to decide if \mathbf{x} belongs to the target class (i.e. $OCGM(\mathbf{x}) > p_0$), the percent of the training class t_0 that represents outliers is used to compute the optimality threshold. Here, we used $t_0=0.05$. If t_0 is too low, possible rare objects or noise might be included in the training set. If t_0 is too big, an entire class of objects (say, shadows) might be excluded.

The OC nearest neighbor (denoted as NN data description, NNdd,) is defined as:

$$OCNN(\mathbf{x}) = \frac{NN(\mathbf{x})}{NN(NN(\mathbf{x}))},$$

where $NN(\mathbf{x})$ is the nearest neighbor of \mathbf{x} .

As opposed to the regular SVM that separates two classes in the feature space by a hyper plane, OCSVM (denoted as support vector data description) surrounds the target class in the feature space by a hyper-sphere. Formally, we need to minimize:

$$R^2 + C \sum_i \xi_i,$$

where ξ_i are slack variables, R is the radius of the hyper-sphere and C is a constant, with the constraints that the objects be in a sphere of radius R :

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \xi_i,$$

where \mathbf{a} is the center of the sphere. In the above formulation, \mathbf{a} and R are computed such that t_0 of the training set objects will lay outside the sphere.

RESULTS

Dataset description

The experiments shown here were performed on an IR video sequence obtained on a 1 mile long country road at an US Army test site. The IR images were obtained using a long wave IR (8-12 μm) camera mounted in front of the vehicle. The video sequence consisted in 1922 frames and had 13 surface road hazards and 3 buried ones. Each road hazard was marked by a square aluminum fiducial (1ft \times 1ft). Although the fiducial were possible "abnormal" objects, we counted the hits that they produced as false alarms.

In order to score our algorithm we marked the abnormal objects in 113 frames (ground truth). Only one target was marked in each frame, even if others were visible farther along the road. We used these frames for computing the receiver-operator curves shown below as follows. A specific target appears (was marked) in 4-6 frames. An abnormal area found around in a window of 40 by 10 centered at target location in any of the ground truth frames was declared a "hit". Any other abnormal objects found in the ground truth frames were declared false alarms.

OCC training

In order to analyze the properties of the three OCC under consideration we extracted about 5000 corners from the first 200 frames of the sequence. In the beginning of the video sequence no targets were visible. The extracted corners belonged to "normal" road side objects such as rocks, bushes, trees or shadows. To train the OCCs we used 4000 corners extracted at random from the 5000 available. Although we don't use the "abnormal" class during the training process, we need it in the testing process. Extracting a large amount of corners from "abnormal" objects is tedious (although we might consider it in future research). Instead, we obtained the second class by randomly permuting the features of the background ("normal") objects. The ROC curves obtained in this fashion for the three OCCs considered are shown in Figure 12.

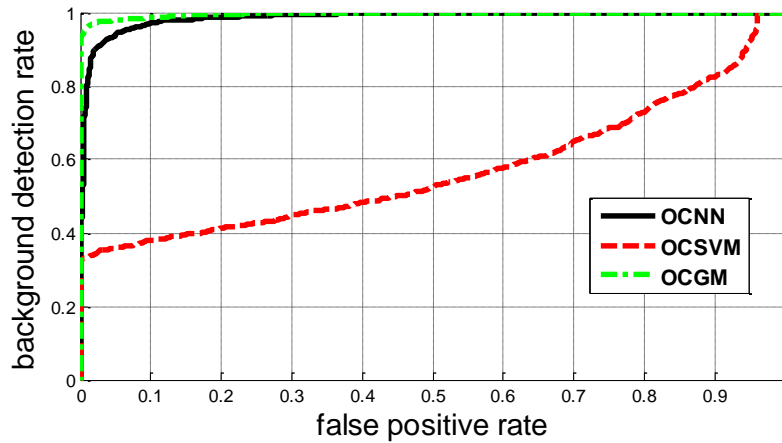


Figure12. Comparison of three OCCs for corners extracted from the first 300 frames.

In Figure 12, the best performance was obtained using OCGM (AROC=0.991). Surprisingly, OCSVM did not perform well (AROC=0.57) although several kernels and rejection fractions t_0 were tried. Consequently, in the following experiment we used OCGM.

COADA detection performance on the available dataset

After we decided on the OCC, we tested several parameters of the corner detection algorithm. The test was performed by running the COADA algorithm (in fact only the corner detection and OCC classification part, without temporal fusion) on the 113 testing frames that had a ground truth ("abnormal") position marked. The scoring procedure was described at the beginning of this section and the results are shown in Figure 13.

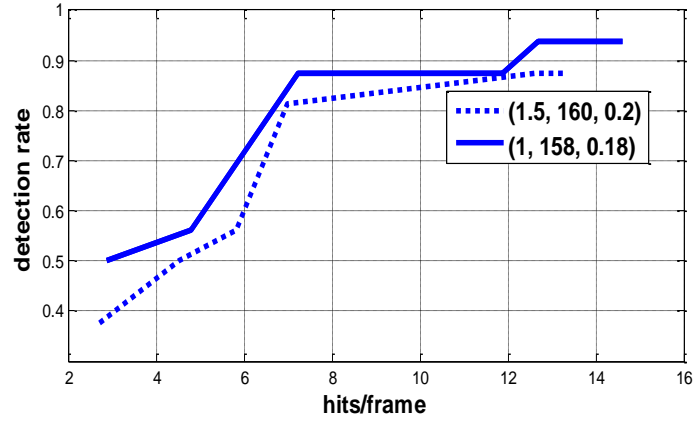


Figure 13. COADA performance for different corner parameters.

As we see from figure 13, the detection algorithm is somewhat sensitive to the parameters of the corner detection algorithm. A slight variation in the object shape (from 1.5 to 1), corner angle (160 to 158) and high Canny threshold (0.2 to 0.18) lead to a sensible better performance. Although the detection performance was acceptable (around 90%, with all surface hazards and even two buried ones discovered - see the continuous line ROC) the number of cues per frame was rather high (around 13). One category of objects that contribute the most to false cues in our case, is bushes. However, the corner hits in a bush have a somewhat random aspect that can be eliminated by temporal fusion.

COADA with temporal fusion performance

We run COADA algorithm on the entire sequence (1922 frames) in order to enable the temporal fusion process. The detection was scored only in the 113 frames where ground truth was available. In the frames without ground truth, all hits were counted as false alarms. This scoring procedure is somewhat pessimistic but not too far from reality due to the prevalence of background objects over road hazards. The resulting performance is shown in Figure 14.

We see that the temporal fusion reduced the false cues by about 40 times with just a slight decrease in detection (no buried targets were not detected in this case). In other words, at a frame rate of 3 images/second, we achieve about 80% detection with a false cue each second.

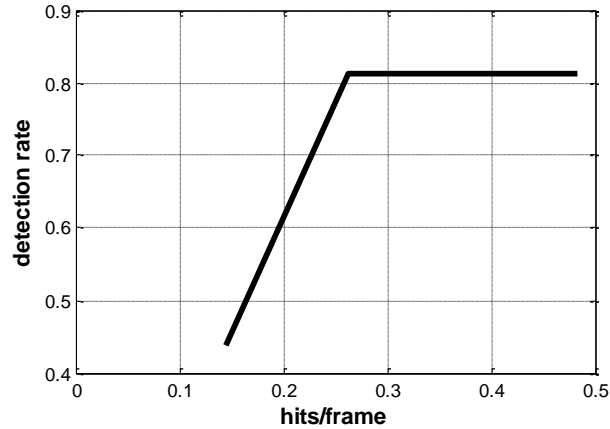


Figure 14. The performance of the COADA algorithm using temporal fusion.

Locally-Adaptive Detection Algorithm for Forward-Looking Ground-Penetrating Radar

The FLGPR images we present here were collected by a system called ALARIC. This system is an FLGPR system that is composed of a physical array of sixteen receivers and two transmitters. In the past decade, FLGPR systems have primarily used their physical arrays (aperture) as well as their radar bandwidth for imaging (resolution); conventional backprojection or time domain correlation imaging has been used for this purpose. Those FLGPR systems rarely tried to exploit imaging information that is created by the motion of the platform. The ground-based FLGPR community has referred to imaging methods that leverage platform motion as multi-look imaging. Though in the airborne radar community, this is better known as *synthetic aperture radar* (SAR) imaging. SAR has been shown to be an effective tool for airborne *intelligence, surveillance and reconnaissance* (ISR) applications.

The ALARIC system is equipped with an accurate GPS system. As a result, we are capable of processing both physical and synthetic aperture imaging even when the platform moves along a nonlinear path with variations in its heading. To create the FLGPR images we use a nonlinear processing technique called *Adaptive Multi-Transceiver Imaging*. This method exploits a measure of similarity among the 32 T/R images which adaptively suppresses artifacts such as sidelobes and aliasing ghosts.

Figure 15 illustrates our proposed explosive-hazard detection algorithm. The sensor fusion with the camera-based sensor is described above. Here, we focus on the locally-adaptive threshold prescreener and the spectrum-feature one-class classifier. We first propose a locally-adaptive detection algorithm. This algorithm builds upon the prescreener that we previously developed. Unlike a conventional threshold-based detector, our algorithm detects local-maxima by applying an adaptive threshold that is sensitive to local noise levels. Test results show that this method reduces the number of FAs by 75%, as compared to a hard threshold-based method, at a probability of detection of 94%. The second algorithm we propose is a classifier that rejects FAs by characterizing the spatial spectrum of FAs. At each alarm-location we compute a 50-bin windowed *fast Fourier transform* (FFT) of the real-part of the FLGPR image. We then train a one-class classifier on these spectrum-based features. We show that we can train a generalized classifier, which is effective at reducing the number of FAs in both training data and test data. Our final results show that we can achieve an approximate FA rate of 0.03 FA/m² at a >90% probability of detection.

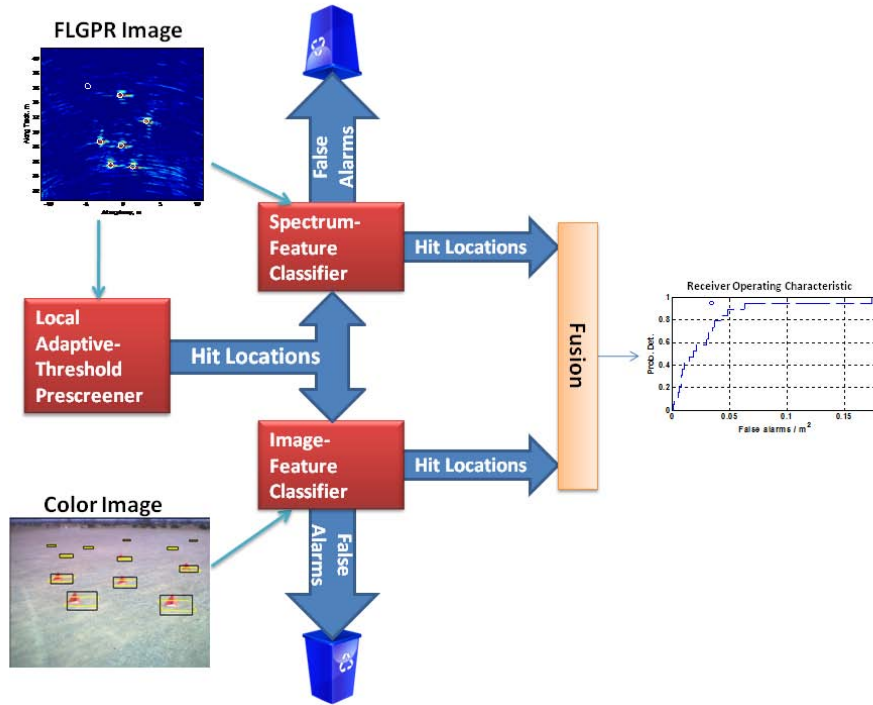


Fig. 15. Block diagram of our forward-looking explosive hazards detection algorithms.

Locally-Adaptive Threshold Detection Algorithm

The FLGPR images are created for an area -11m to 11m in the cross-range direction (although, in practice, only a sub-region of this is used in our detection algorithms), where negative numbers indicate to the left of the vehicle. Coherent integration of radar scans is performed in an area 9m to 25m in front of the vehicle. The pixel-resolution of the FLGPR image is 0.05m x 0.05m. The nominal center frequency is 1.2GHz and the bandwidth is 1.5GHz. We chose a detection region 9m wide. If the targets are on the left side of the road (relative to the vehicle) this region is positioned from -7m to +2m; if the targets are on the right side of the road this region is positioned from -2m to +7m. The prescreener algorithm we present here is an extension of this previous work.

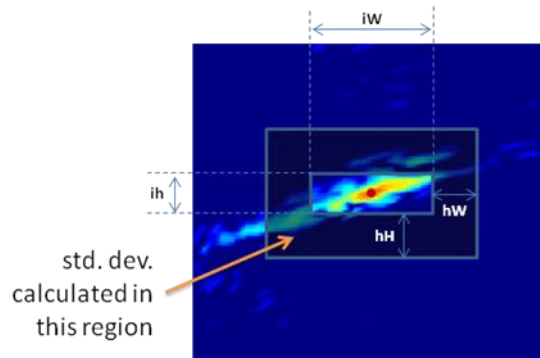


Fig. 16. Local adaptive-threshold prescreener calculates standard deviation in rectangular halo around each radar image pixel.

Detection algorithm

Consider an FLGPR image $G(u, v)$, where u is the cross-range coordinate and v is the down-range coordinate. We first filter G with a locally-adaptive standard deviation filter. This computes the local standard deviation in a variable-size rectangular halo around each pixel. Figure 16 shows the region in which the local standard deviation is calculated. We define this region by the dimensions of the inner rectangle and the width of the outer halo. Each pixel in $G(u, v)$ is divided by the local standard deviation

$$G_f(u, v) = \frac{G(u, v)}{\sigma_{\text{local}}(u, v)}, \forall u, v,$$

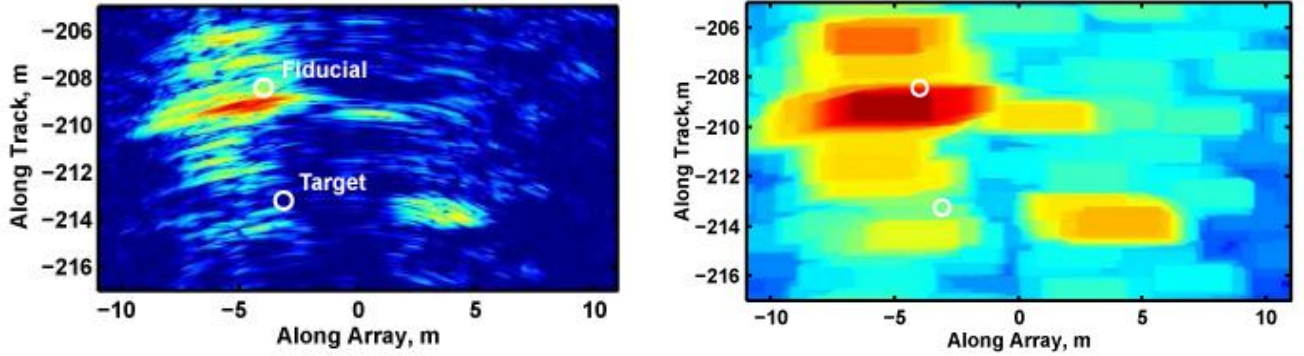
where $\sigma_{\text{local}}(u, v)$ is the standard-deviation of the pixels within the halo region around $G(u, v)$.

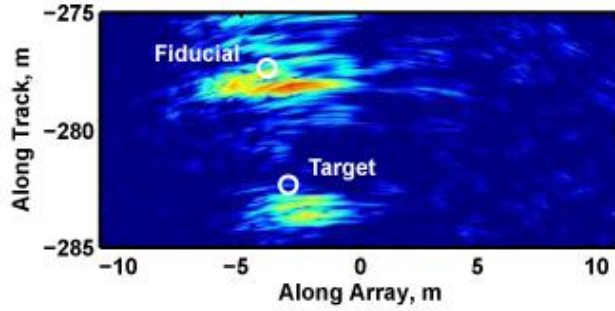
The filtered image is then input to a local-maxima finding algorithm. Our detection method first computes a maximum order-filtered image with a 3m x 1.5m kernel. We denote this order-filtered image as $O_f(u, v)$. Essentially, each pixel in the scan image is replaced by the maximum pixel value within a 3m crossrange by 1.5m downrange rectangle. Figure 17 shows two examples of FLPR images and their associated order-filtered images. As this figure shows, the order-filter reduces the noise-induced artifacts in the image and shows the local maxima as large squares in the image. Alarms are identified by the operation

$$A = \arg_{\{u, v\}} \{G_f(u, v) \geq \min\{O_f(u, v), -60\}\},$$

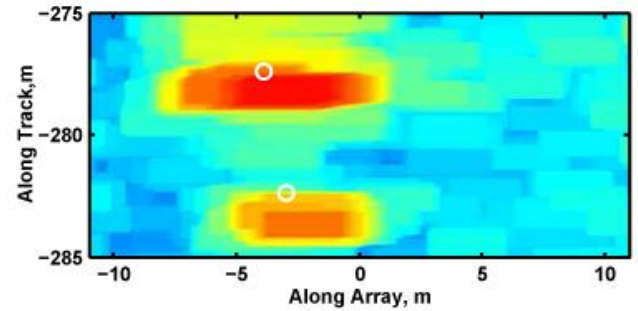
where A is the set of local-maxima locations. The minimum operator prescreens alarm locations that have a very low FLGPR return. We choose a value of -60dB for this threshold as this only eliminates alarms with the lowest of confidence (note that the minimum value in the color scale in Fig. X is -8dB). This prescreening threshold merely minimizes the computational cost of the subsequent algorithms by reducing the number alarms to a manageable number. We also augment each alarm location (u, v) in A with the value of the FLGPR image pixel at that location, which we denote as $G_f(A)$. This pixel value is, in effect, the confidence of the alarm – the higher the pixel value (FLGPR return), the higher the confidence. Figure 4 shows the associated alarm locations of the images shown in Fig. 3.

As Figs.17 and 18 show, there were fiducials (markers) placed near the target locations in the tests. We identified fiducial hits and removed them from our ROC calculations. The fiducial hits in Fig.18 are denoted by the '+' symbol. Note that our method for identifying fiducial hits is not perfect, but adding or subtracting one alarm location only negligibly affects the overall ROC results.





(a) Local standard-deviation filtered images



(b) Maximum order-filtered images

Fig. 17. Maximum order-filtered images of FLGPR images – target locations indicated by white circles.

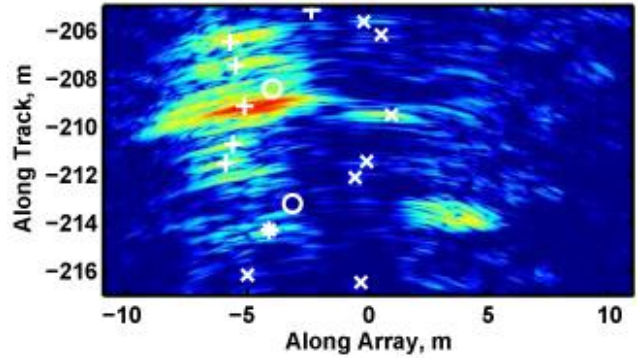
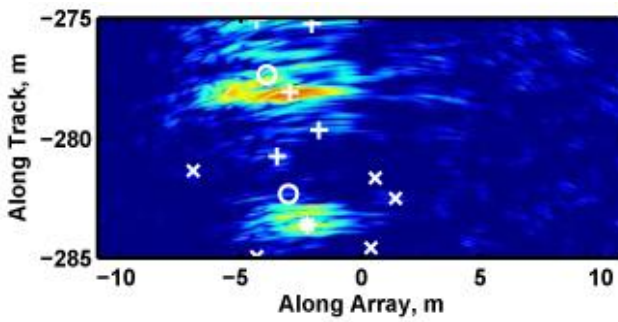


Fig. 18. Alarm locations for example images in test run 188. x indicates FA, + indicates fiducial alarm, and * indicates a target alarm.

Figure 19 displays the effectiveness of our locally-adaptive threshold detection algorithm. The solid blue line indicates the performance of a non-adaptive conventional threshold detector. As the ROC curve shows, this algorithm detects only 88% of the targets at a FA rate of 0.16 FA/m^2 . The dotted lines indicate the performance of our locally-adaptive algorithm for four different sized windows (see Fig. 2 for an illustration of the window size). The 5×5 , 5×20 window size achieved the best FA rate at a detection probability $> 90\%$. This window size results in a minimum FA rate of 0.045 FA/m^2 at a probability of detection of 94%. We stress that all instances of the locally-adaptive threshold detector were able to achieve a probability of detection of 100% with less than 0.1 FA/m^2 .

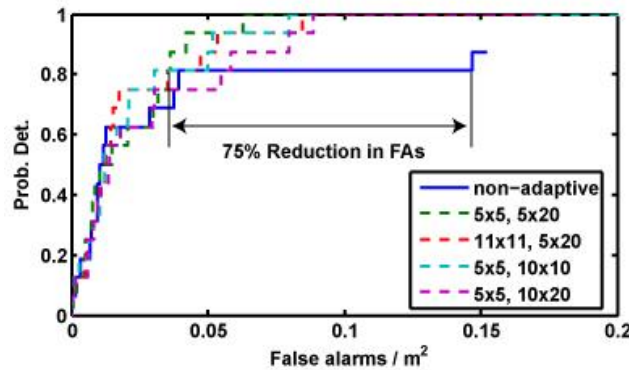


Fig. 19. ROC curve of MUFL prescreener for non-filtered radar image and three different sized locally-adaptive filter halos. The size of the rectangular halo is denoted as $iW \times iH$, $hW \times hH$, as shown in Fig. 16.

Spectrum-based False alarm rejection

Spectrum-feature

A spectrum-based feature is calculated for each FLGPR detection. We first calculate a 50-bin windowed FFT of the row of pixels centered at the detection location

$$X_f(A) = |FFT(W * G_f(A)\{-25:24\})|,$$

where W is a 50-point Hamming window and $X_f(A)$ is the magnitude of the windowed-spectrum of $G_f(A)\{-25:24\}$, the 50-point horizontal slice of the FLGPR image centered at the alarm A . We use the 50-bins of $X_f(A)$ as the features of a one-class classifier that is trained on the FA locations. Essentially, the one-class classifier is a model of the spectrum of the FAs.

One-class classifier

The 50 spectrum-based features and the FLGPR confidence value for each detection are used to classify the detection as either true (an explosive hazard) or false. We train a classifier by first calculating the multivariate normal distribution that best represents the feature values of the false detections for a given set of training data. Hence, the values of the false detections are assumed to be accurately represented by

$$f_X(x_1, \dots, x_{50}) = \frac{1}{(2\pi)^{50/2} |\Sigma|^{0.5}} \exp(-0.5(x - \mu)^T \Sigma^{-1}(x - \mu)),$$

where μ is the mean vector, Σ is the covariance matrix, and $\{x_1, \dots, x_{50}\}$ are the 50 features in $X_f(A)$. We fit the distribution parameters to the training data using the well-known maximum-likelihood estimator.¹⁶ Once we have trained the classifier, we can use the Malanhobis-metric to determine how well a new feature vector X fits the false detection distribution, where this distance is calculated by

$$D(X) = \sqrt{(X - \mu)^T \Sigma^{-1}(X - \mu)}.$$

If the Malanhobis-metric $D(X)$ is large-valued, this indicates that the detection does not fit the false detection distribution and is, most likely, a true detection. Hence, a threshold T must be chosen such that a $D(X) > T$ indicates a true detection and a $D(X) \leq T$ indicates a false detection. The advantage of this method is that the threshold T can be tuned to offer an optimal tradeoff between true and false detections. Also, the distribution is trained on false detection data, of which there are many, rather than true detection data, of which there are few. Furthermore, the true detection features can be drastically different for different types and configurations of the explosive hazards, whereas the false detection features tend to more generalized.

Feature and Threshold Selection

There are a total of 50 spectrum-based features for each FLGPR detection. It is unlikely that all of these features are necessary or effective for training an optimal classifier. Additionally, given a set of features we must choose the threshold T which determines whether an input feature vector is classified as a true or false detection. We use an *exhaustive search* to find the four best features. Earlier, we used a forward sequential search to determine the best N features. However, we have since discovered that with an exhaustive search can be performed relatively quickly and produces more generalized classification results. At each iteration of the exhaustive feature selection, the threshold T is set such that each target in the training data has at least one associated detection. In this manner, the optimal T eliminates the most false detections while maintaining a $P_D = 100\%$. Thus, the exhaustive search determines the four best features and associated classifier parameters, μ , Σ , and T .

Figure 20(a) shows the training results of using the spectrum-based classifier on the alarm locations following the locally-adaptive threshold prescreeener. The training data is Test Run 188. These results show that the classifier is able to reduce the FA rate from 0.045 FA / m² to 0.022 FA / m² – a greater than 50% reduction. We note, however, that these are resubstitution results and represent the best performance that would be expected from this classifier.

RESULTS

Locally-adaptive prescreener results

Figure 20 shows the ROC curves of the locally-adaptive prescreener on test runs 188 and 190. The size of the local standard-deviation filter used was 5×5 , 5×20 (see Fig.16 for an illustration of the filter dimensions), which was the most effective filter size on test run 188 (as shown in Fig. 18). All results shown in this section will use this filter size. On test run 188 our prescreener is able to achieve a minimum FA rate of 0.045 FA/m^2 at 94% probability of detection. On test run 190 the prescreener produces a minimum FA rate of 0.34 FA/m^2 at 90% probability of detection. Figure 20 shows that this prescreener not only effective on the training data (188) but also on the test data (190).

Spectrum-feature classifier results

Figure 21 outlines the FA rejection results for the one-class classifier trained with the spectrum-feature. A confidence threshold was chosen from the training data (test run 188) that resulted in a $>90\%$ classification rate with the least number of FAs. This is shown as the cyan dot in view (a) – this is the expected performance using just the locally-adaptive prescreener. As Fig. 21(a) illustrates, the FA rate of the locally-adaptive prescreener at 94% probability of detection is 0.045 FA/m^2 . The red dot in view (a) shows the FA rate after the spectrum-feature classifier is applied. As this shows, the FA rate was reduced by $>50\%$ to 0.022 FA/m^2 .

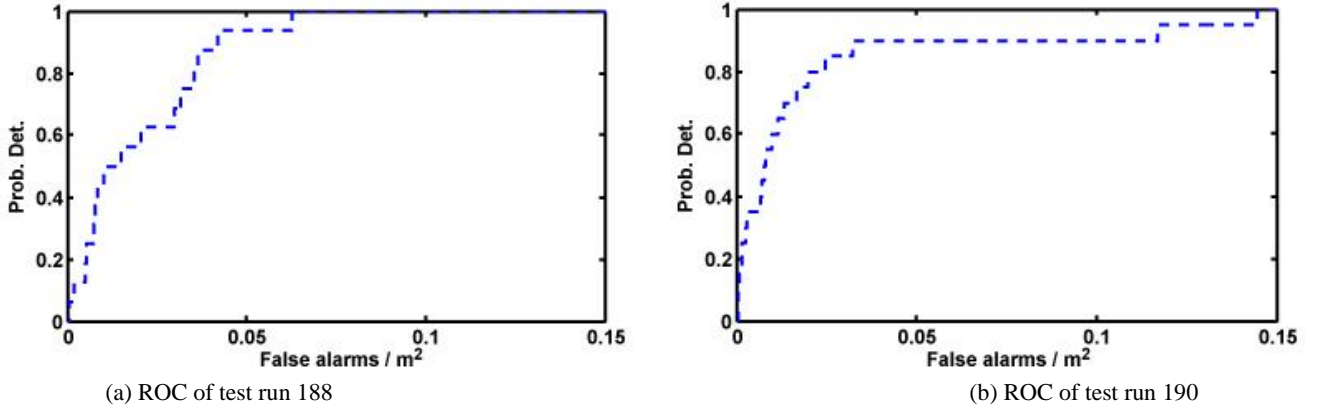


Fig. 20. Results of locally-adaptive threshold detector on test runs 188 and 190.

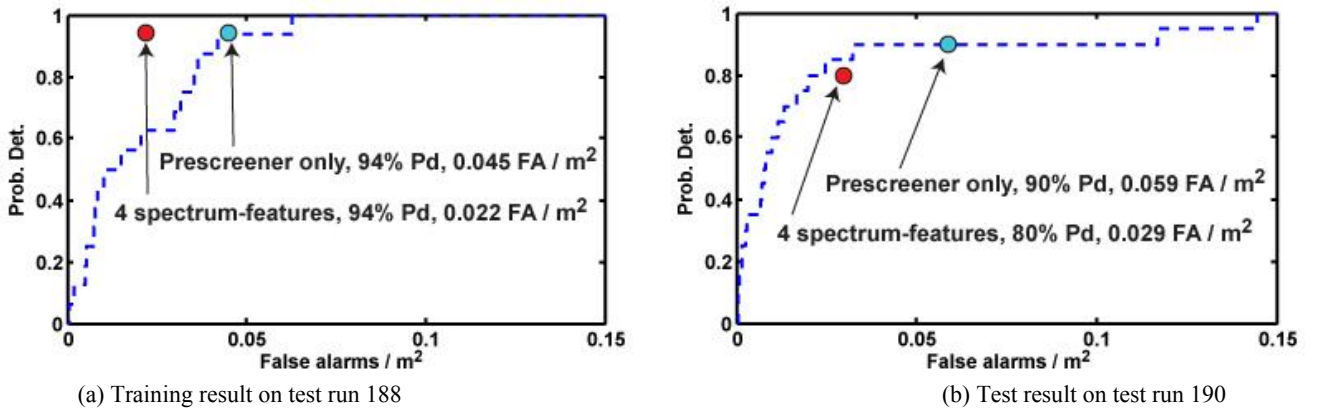


Fig.21. Training and test results of one-class classifier with 4 spectrum-based features – bins [21,27,30,50] of FFT. Feature selection based on best training results.

The same confidence threshold was then applied to test run 190. View (b) shows that the locally-adaptive prescreener, with the threshold chosen from the training data in view (a), results in 90% probability of detection with 0.059 FA/m^2 (shown by the cyan dot). If we apply the trained spectrum-feature classifier to test run 190, we only achieve a

probability of detection of 80% with a FA rate of 0.029 FA/m². This is clearly undesirable as the probability of detection is reduced. However, recall that only 4 of the 50 spectrum features were used in the training of the classifier. Thus, we examined other combinations (of 4 features) of the 50 spectrum features to identify features that would better generalize across the two data sets.

Figure 22 shows the results of the spectrum-feature classifier using a different set of 4 features. The 4 features were chosen that resulted in the best average training and test performance. Note that the classifier is still trained only on the training lane (188). However, by selecting a different set of features we were able to train a classifier that has a more generalized effectiveness. View (a) shows that using bins [15, 17, 30, 39] of the FFT results in a 94% probability of detection with 0.026 FA/m² on the lane 188 – in the pattern recognition community these are often called resubstitution results. In view (b), we show the results of the trained classifier on lane 190, the test data. With these 4 features, the classifier produces a 90% probability of detection with 0.034 FA/m². Although the FA rates in both the training and test data are slightly higher than those shown in Fig. 21, in contrast the test lane performance is much better as the probability of detection is maintained at 90%. These results are promising as this shows that we can build a generalized spectrum-feature classifier that significantly reduces the number of FAs in both training and test data.

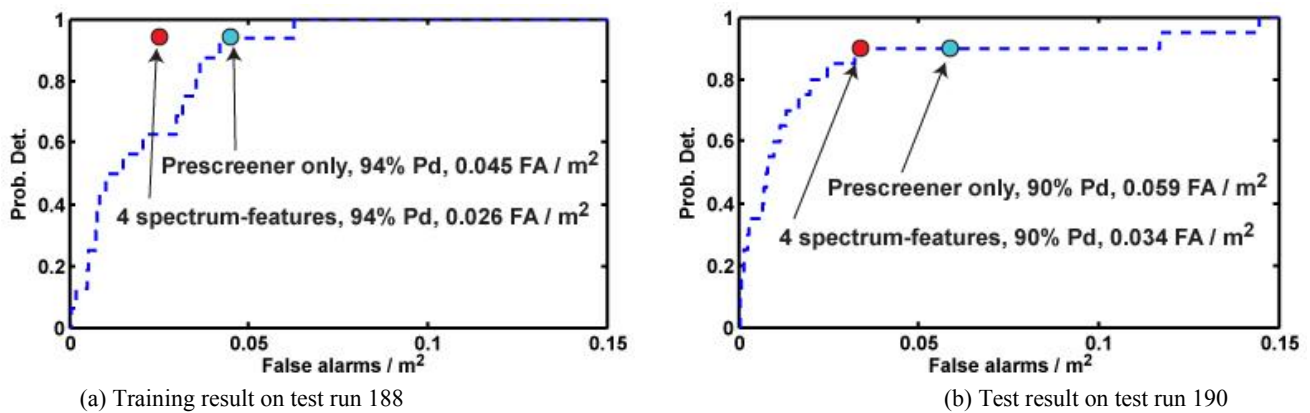


Fig. 22. Training and test results of one-class classifier with 4 spectrum-based features – bins [15,17,30,39] of FFT. Feature selection based on best average training and test results. This feature selection method results in a more generalized classifier.

Improved Detection and False Alarm Rejection Using FLGPR and Color Imagery in a Forward-Looking System

CAMERA-BASED FALSE ALARM REJECTION

Using the methods described above, we are able to find the areas in the camera images that correspond to each FLGPR detection. Hence, we can use the information in the IR images to classify the types of detections from the FLGPR, assuming that the image pixels corresponding to a false detection (e.g. bushes, rocks, garbage, etc.) are different from the pixels corresponding to an explosive hazard. The camera used on the NVESD system is a 1024x768 visual-spectrum color camera. The camera is aimed forward such as to image the same portion of the scene at which the FLGPR is radiating. Figure 23 shows an example of one of these images. For this paper, we focused on developing a robust and simple method for using the camera images to classify FLGPR detections as either true or false detections.



Fig. 23. Example of camera image taken by system.

Color Feature Extraction

Each FLGPR detection can be projected into a camera pixel location (assuming that the detection is within the camera field-of-view). Generally, there are multiple frames, between 15 and 30, for each FLGPR detection. The distance to the detection location differs in each frame, and, therefore, the number of pixels that targets comprise in a corresponding camera image differs. We are interested in examining a fixed area, in meters, around each detection location; thus, an adaptive-sized window around each detection in the image is selected. The projection matrix PR allows us to compute the size of each image pixel, in meters, by using the inverse transformation from pixel positions to camera reference frame coordinates. Hence, it is possible to determine the appropriate window size to use for each image position, which corresponds to a chosen real world distance. We use a window size corresponding to a side length of one meter in the horizontal direction (cross-range) and two meters in the vertical direction (down-range), as we discovered that this is large enough to contain all targets present in our data. We denote these sub-images as W .

We calculate a set of features from the pixels in the windows corresponding to each FLGPR detection. First, the intensity, local standard deviation, Laplacian, and Sobel images are calculated. The Laplacian is calculated using the convolution kernel

$$\begin{bmatrix} 1/3 & 2/3 & 1/3 \\ 2/3 & -4/3 & 2/3 \\ 1/3 & 2/3 & 1/3 \end{bmatrix}.$$

The local standard deviation is calculated in a 5x5 window around each pixel. The Sobel image is calculated as

$$S = (W * S_x)^2 + (W * S_y)^2,$$

where $*$ indicates convolution and the squares are calculated element-wise. We use the standard Sobel gradient operators, denoted as S_x and S_y . We also create three other images, one each of the red, blue, and green channels of the image.

The set of features calculated on the target detections in each of the seven images (intensity, local standard deviation, Laplacian, Sobel, red, green, and blue) are the average, minimum, maximum, median, standard deviation, skewness, and kurtosis. For example, consider the red-channel image. The seven features corresponding to a sub-image W would be the average red pixel-value in W , the minimum red pixel-value in W , the maximum red pixel-value in W , etc. In total, 49 features are calculated from each window W , which is the sub-image where an FLGPR detection is visible. Recall that each detection location can appear in multiple images (usually 15-30); thus, each detection is represented by 15 to 30 sets of the 49 camera-based features. The median of these 15-30 sets of features is calculated so that each detection is represented, finally, by 49 aggregate feature values. We have experimented with other feature aggregation methods,

including mean (both conventional and alpha-trimmed), min, and max, and we discovered that median was the most effective aggregation operator for combining the features from the multiple camera frames. In the future we hope to examine methods by which all sets of features can be used.

We then train a one-class classifier to reject FAs based on the 49 aggregate features.

One-class classifier

The 49 camera-based features and the FLGPR confidence value for each detection are used to classify the detection as either true (an explosive hazard) or false. We train a classifier by first calculating the multivariate normal distribution that best represents the feature values of the false detections for a given set of training data. Hence, the values of the false detections are assumed to be accurately represented by

$$f_X(x_1, \dots, x_{49}) = \frac{1}{(2\pi)^{\frac{49}{2}} |\Sigma|^{0.5}} \exp(-0.5(x - \mu)^T \Sigma^{-1}(x - \mu)),$$

where μ is the mean vector and Σ is the covariance matrix. We fit the distribution parameters to the training data using the well-known maximum-likelihood estimator. Once we have trained the classifier, we can use the Mahalanobis-metric to determine how well a new feature vector X fits the false detection distribution, where this distance is calculated by

$$D(X) = \sqrt{(X - \mu)^T \Sigma^{-1}(X - \mu)}.$$

If the Mahalanobis-metric $D(X)$ is large-valued, this indicates that the detection does not fit the false detection distribution and is, most likely, a true detection. Hence, a threshold T must be chosen such that a $D(X) > T$ indicates a true detection and a $D(X) \leq T$ indicates a false detection. The advantage of this method is that the threshold T can be tuned to offer an optimal tradeoff between true and false detections. Also, the distribution is trained on false detection data, of which there are many, rather than true detection data, of which there are few. Furthermore, the true detection features can be drastically different for different types and configurations of the explosive hazards, whereas the false detection features tend to more generalized. In practice, if one is using $D(X)$ to produce a threshold detector, then the square-root does not need to be included.

Feature and Threshold Selection

There are a total of 49 camera-based features for each FLGPR detection. It is unlikely that all of these features are necessary or effective for training an optimal classifier. Additionally, given a set of features we must choose the threshold T which determines whether an input feature vector is classified as a true or false detection. We use an *exhaustive search* to find the four best features. We have discovered that an exhaustive search can be performed relatively quickly and produces more generalized classification results. At each iteration of the exhaustive feature selection, the threshold T is set such that each target in the training data has at least one associated detection. In this manner, the optimal T eliminates the most false detections while maintaining a $P_D > 90\%$. Thus, the exhaustive search determines the four best features and associated classifier parameters, μ , Σ , and T .

RESULTS

Spectrum-feature classifier test results

Figure 24 outlines the FA rejection results for the one-class classifier trained with the spectrum features. A confidence threshold was chosen from the training data that resulted in a $>90\%$ classification rate with the least number of FAs. This is shown as the blue dot in view (a) – this is the expected performance using just the locally-adaptive prescreener. As this figure shows, the expected FA rate at 95% probability of detection is 0.06 FA/m². The red dot in view (a) shows the FA rate after the spectrum-feature classifier is used. As this shows, the FA rate was reduced by 33% to 0.04 FA/m².

The same confidence threshold was then applied to Test Run B. View (b) shows that the locally-adaptive prescreener, with the threshold chosen from the training results in view (a), results in 90% probability of detection with 0.11 FA/m² (shown by the blue dot). If we apply the trained spectrum-feature classifier to Test Run B, we only achieve a

probability of detection of 75% with a FA rate of 0.06 FA/m². This is clearly undesirable. However, recall that we use only 4 of the 50 spectrum features in the training of the classifier. Thus, we examined other combinations (of 4 features) of the 50 spectrum features to see if we could find features that would better generalize across the data sets.

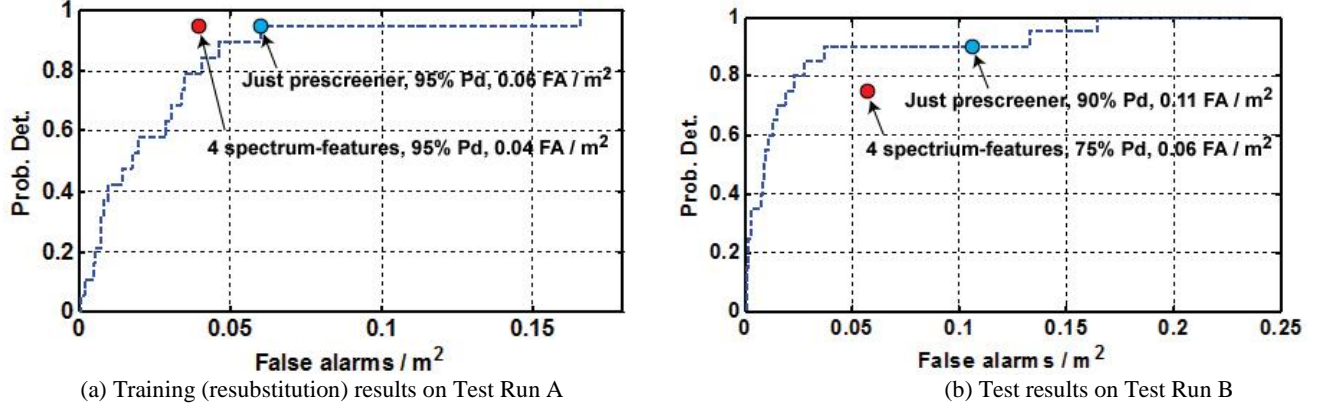


Fig. 24. Training and testing results of one-class classifier with 4 spectrum-based features – bins [23,32,33,50] of FFT. Feature selection based on best training (resubstitution) results.

In a second experiment, we examined other sets of spectrum-features to determine if we could find a set of 4 features that would result in better generalized performance. Figure 24 illustrates the results of this experiment. We first trained a spectrum-feature classifier on Test Lane A (the training lane) for all possible sets of 4 spectrum-based features. We then examined the resulting performance on Test Lane B (the testing lane). View (b) shows the resulting detection characteristics for the classifier using bins [22, 29, 39, 42] of the spatial FFT. As this plot shows, by using these features the FA rate on the test lane was reduced from 0.11 FA/m² to 0.06 FA/m² while maintaining a 90% probability of detection. View (a) shows that the training lane performance is slightly degraded as compared to the results in Fig. 21(a); however, we stress that there is still a 15% reduction in FAs. The results shown in Fig. 25 are promising as this shows that by choosing a different set of features, we can train a classifier that performs better for both the training data and the testing data.

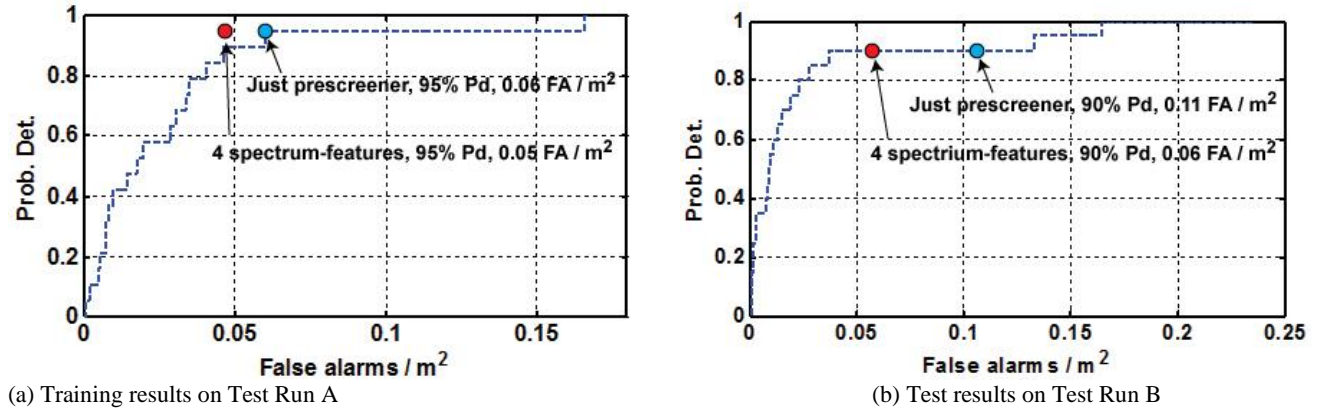


Fig. 25. Test results of one-class classifier with 4 spectrum-based features – bins [22,29,39,42] of FFT. Feature selection based on best test results. This feature selection method results in a more generalized classifier.

3.2 Image-feature classifier test results

Figure 26 illustrates the performance of the image-feature classifier. The red dot in view(a) indicates the performance using the set of 4 camera-based features that minimize the FA rate while maintaining at least 90% probability of detection on the training data, Test Run A. The 4 features selected by our exhaustive search were skewness of the pixel intensity, the minimum of the Laplacian, the mean of the Laplacian, and the median of the Laplacian. View (b) shows the resulting performance of the trained image-feature classifier on the test data, Test Run B. As this plot shows, the probability of detection was not reduced; however, the FA rate was negligibly reduced. Note that the results in this section do not

include the spectrum-feature classifier described in Section 3.1. In Section 3.3 we specifically discuss fusing the two classifiers.

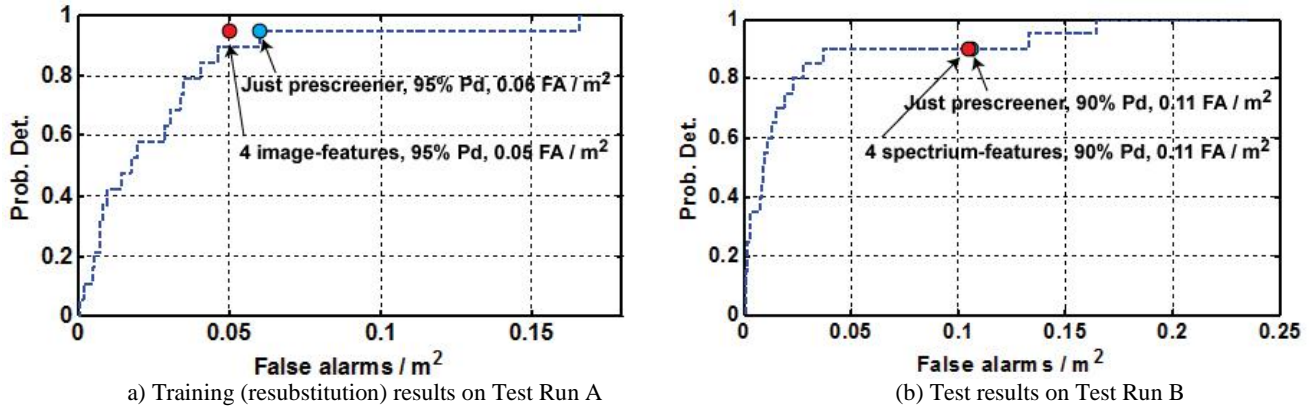


Fig. 26. Training and testing results of one-class classifier with 4 image-based features – skewness(intensity), minimum(Laplacian), mean(Laplacian), median(Laplacian). Feature selection based on best training (resubstitution) results.

We then ran a second experiment in which we examined other sets of 4 image features, with the intention of finding a set that better generalized. Thus, we trained the classifier on all possible sets of 4 image features from the training data, Test Run A, and then examined the performance of these classifiers on Test Run B. Figure 27 shows that using the skewness of the pixel intensity, the skewness of the Laplacian, the median of the local standard deviation, and the minimum of the red channel results in a more generalized classifier. The FA rate on the test data was reduced from 0.11 FA/m² to 0.08 FA/m² at 90% probability of detection. Notice, however, that the FA rate in the training data was only slightly reduced. However, we believe that this method of selecting the features results in a more generalized classifier, which is essential in an operational system.

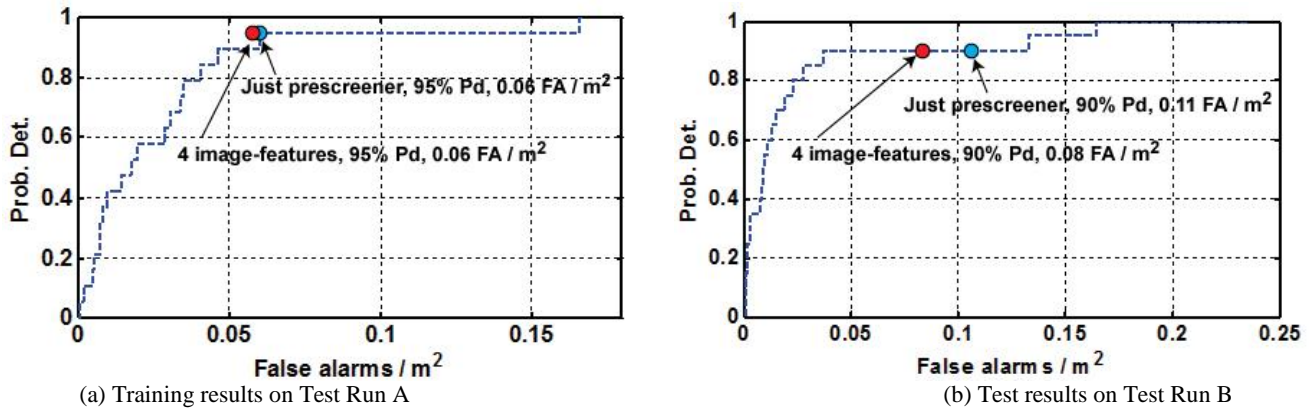


Fig.27. Training and testing results of one-class classifier with 4 image-based features – skewness(intensity), skewness(laplacian), median(local standard deviation), minimum(red channel). Feature selection based on best test results.

Fusion test results

We now show the performance of the system when these two classifiers are fused. The first step in our detection algorithm is to apply the locally-adaptive threshold detector.¹¹ The ROC curve of this detector is shown as the blue dotted line in all the figures in this section. Thus, we first choose a threshold that gives the least number of FAs with at least 90% probability of detection. This is shown as the blue dots in Fig. 28. Second, we fuse the spectrum- and image-feature classifiers using a logical OR. If either classifier determines that an alarm is a FA then the fused result is a FA.

Figure 28 shows the results of our fusion experiment. View (a) shows the resulting FA rate on the training data and view (b) shows the resulting FA rate on the testing data. For these results, we used the set of features that resulted in the best

generalized classifier performance – these features are listed in the captions of Figs. 25 and 27. As Fig. 28 shows, the fusion of the spectrum- and image-features classifiers causes significant reduction in FAs in both the training data and the testing data. The training data FA rate was reduced from 0.06 FA/m² to 0.03 FA/m², a 50% reduction, while maintaining a 95% probability of detection. The FA rate in the test data was reduced from 0.11 FA/m² to 0.05 FA/m² while maintaining a 90% probability of detection. These results show that our FA rejection method is very effective.

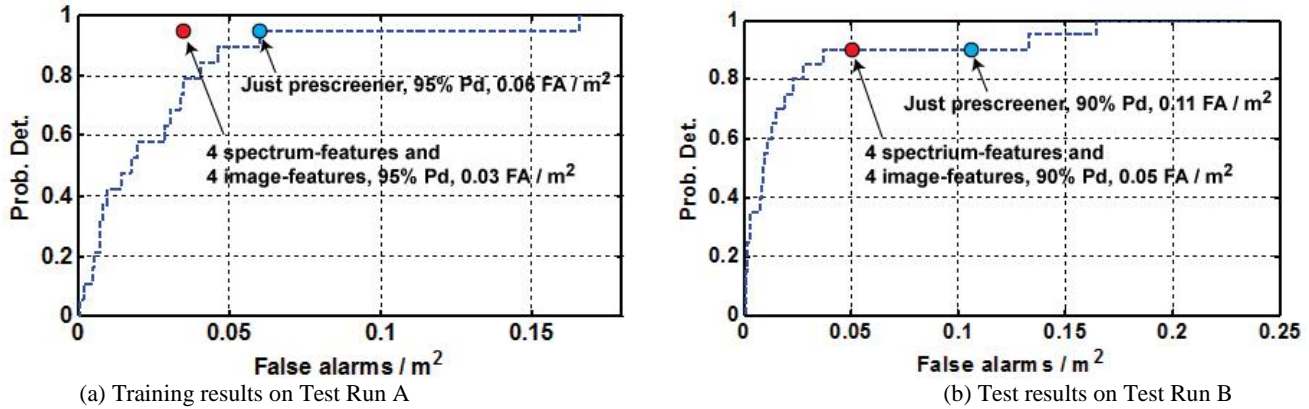


Fig. 28. Test results and training results of fusion of spectrum- and image-based false alarm rejection methods. Feature selection based on best test results

Feature Extraction in Multi-Modal Forward Looking Imagery

Our overall research project can be characterized as one where computer algorithms attempt to locate instances of specific objects within a large data set of images; or, given any point on an image, to return the probability that a specific object is local within some radius. The process involves characterizing images of certain types of objects. More specifically, for multiple sets of color images (frames) in which a consistent time interval separates every consecutive image in a set, the objective of this project is to develop a means for collecting, storing, and accessing images of specific objects (sub-image) extracted from the frames (super-images); to collect, calculate, and store information describing each sub-image; and to associate sets of temporally linked sub-images, which move through super-image space with respect to time (with respect to frame index).

The process involves characterizing images of certain types of objects. To aid in analysis, the sub-images and associated information are pre-extracted and sorted in a database. This allows specific sets of data to be analyzed at once while excluding other sets of data. It also reduces the computer processing time necessary to locate and analyze the data.

The database stores structured information pertaining to multiple sets of temporally linked sub-images. It is a collection of two object-oriented classes: *Sequence* and *Datanode*. Each instance of the *Sequence* class contains data regarding exactly one specific object over some range of frames. It holds information about the object's type and the data set in which it can be found, as well as an array of *Datanodes*. Each instance of the *Datanode* class contains data regarding exactly one frame of the specific object. It holds information about the file in which the sub-image can be found and its coordinates on the super-image.

The database does not directly store any image data. Image data is stored in a different directory within an umbrella directory. This approach allows loading the database without the overhead of loading hundreds of megabytes of images. This also greatly improves the efficiency of analyzing a partial data set and for developing new image features with which to characterize specific object types.

We developed the MATLAB application *Sequence Extraction Graphic User Interface* (SEG), which is shown in Figure 29, to conveniently collect data to populate the database. This application can display the sequential set of super-

images from which to extract the data. The user can label the object as a certain type and select the positions and number of instances of extracted sub images. The SEG application organizes the multiple sub-images of a single object and stores them in a Sequence, which is then added to the database.

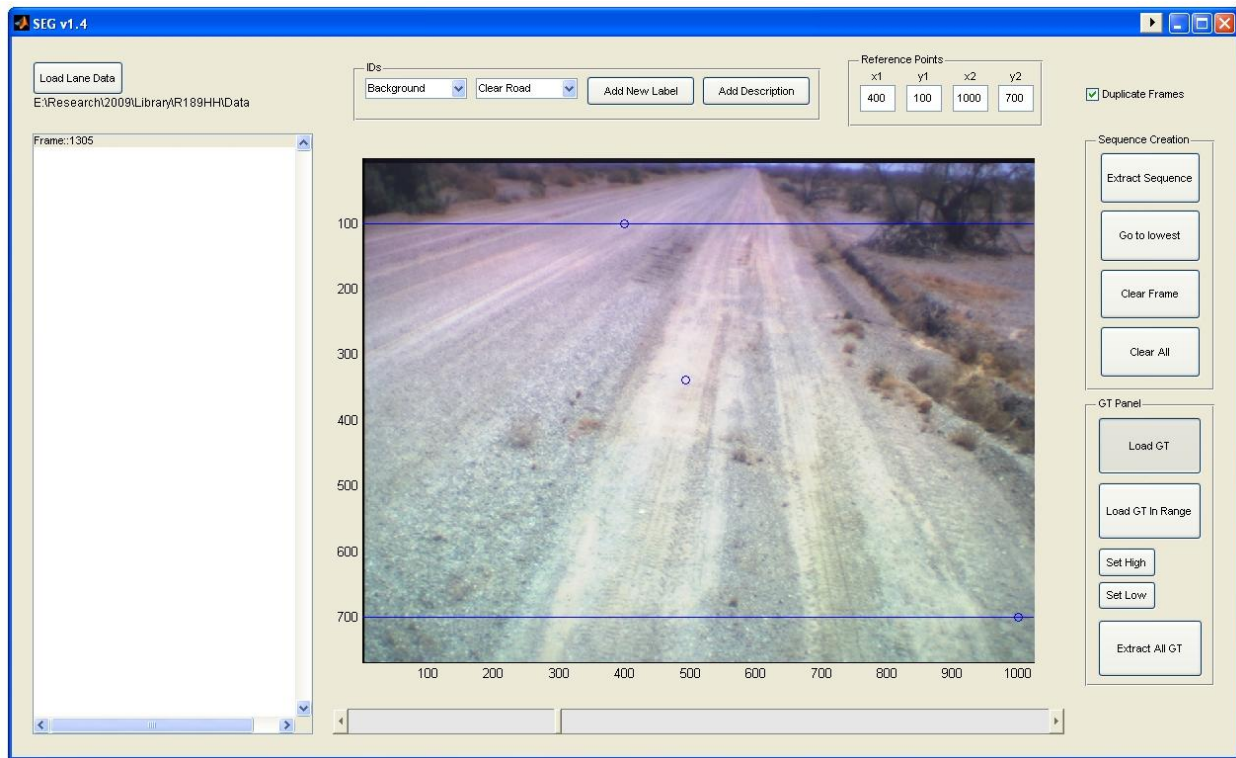


Figure 29. Example rendering of the SEG MATLAB application

The SEG was developed to help create a database of image sequences. In its current version, SEG requires only two files to run. SEG needs the GPS locations of the cart for a particular image and a lane info file. The lane info file contains identification information that is stored in the database along with any information that is extracted from the images. Once the necessary files are loaded, data for a particular object can be extracted based on mouse clicks from the user or a ground truth file with northing and easting coordinates.

By default, the ground truth file only shows object locations, but this file can also be used to extract a single object or the entire lane of objects. Extracting information based on the ground truth is an automated process and allows the user to quickly enter hundreds of sequences into the database with minimal trouble. SEG's ability to label sequences and add descriptions before the sequence is added in the database makes it easy to sort through the database to find what you are looking for.

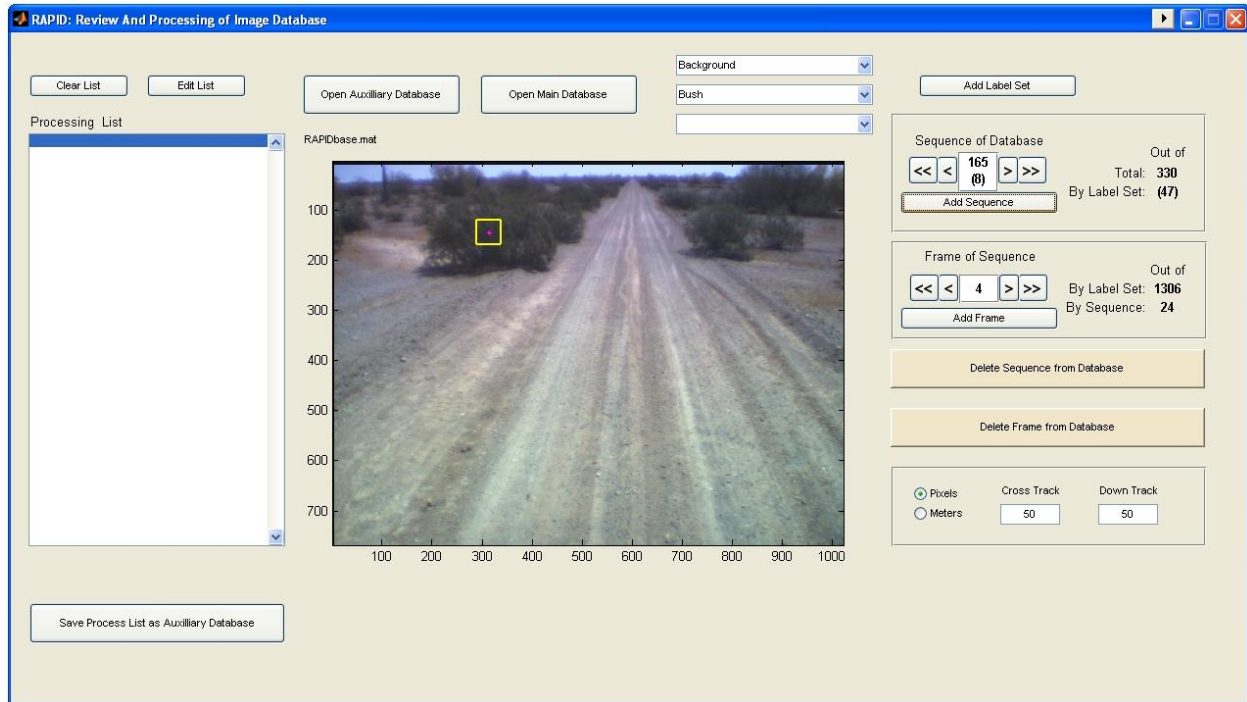


Figure 30: Example rendering of the RAPID MATLAB application.

To conveniently review the content of the database, we developed the MATLAB application *Review and Processing of Image Database* (RAPID), see Figure 30. This application can display any super-image or sub-image of a target. The user can browse the database by data set and/or object type. Sequences or individual Datanodes can be permanently removed from the database. The user can also make a list of interesting data and save it as a separate, auxiliary database (e.g. all sub-images of green bushes). The end result of the RAPID application is a refined database set with a common format, which can be efficiently analyzed using additional MATLAB tools.

Each hit instance appears in a sequence of typically 20 to 30 consecutive video frames. SEG constructs a set of statistical feature vectors for each video sequence corresponding to a hit instance. Each vector contains statistical information relating to a 100 x 100 set of pixels centered on each hit (approximately 2m down range and 1m cross-range). See Figure 31. Seven statistics are computed for each hit instance: (1) image intensity, (2) Laplacian of intensity, (3) Sobel edge feature of intensity, (4) Local standard deviation of intensity, (5) red channel, (6) green channel, and (7) blue channel. The following attributes are computed for each statistic: (1) max, (2) min, (3) mean, (4) median, (5) standard deviation, (6) skewness and (7) kurtosis. Thus, each vector associated with a hit instance has 49 components.

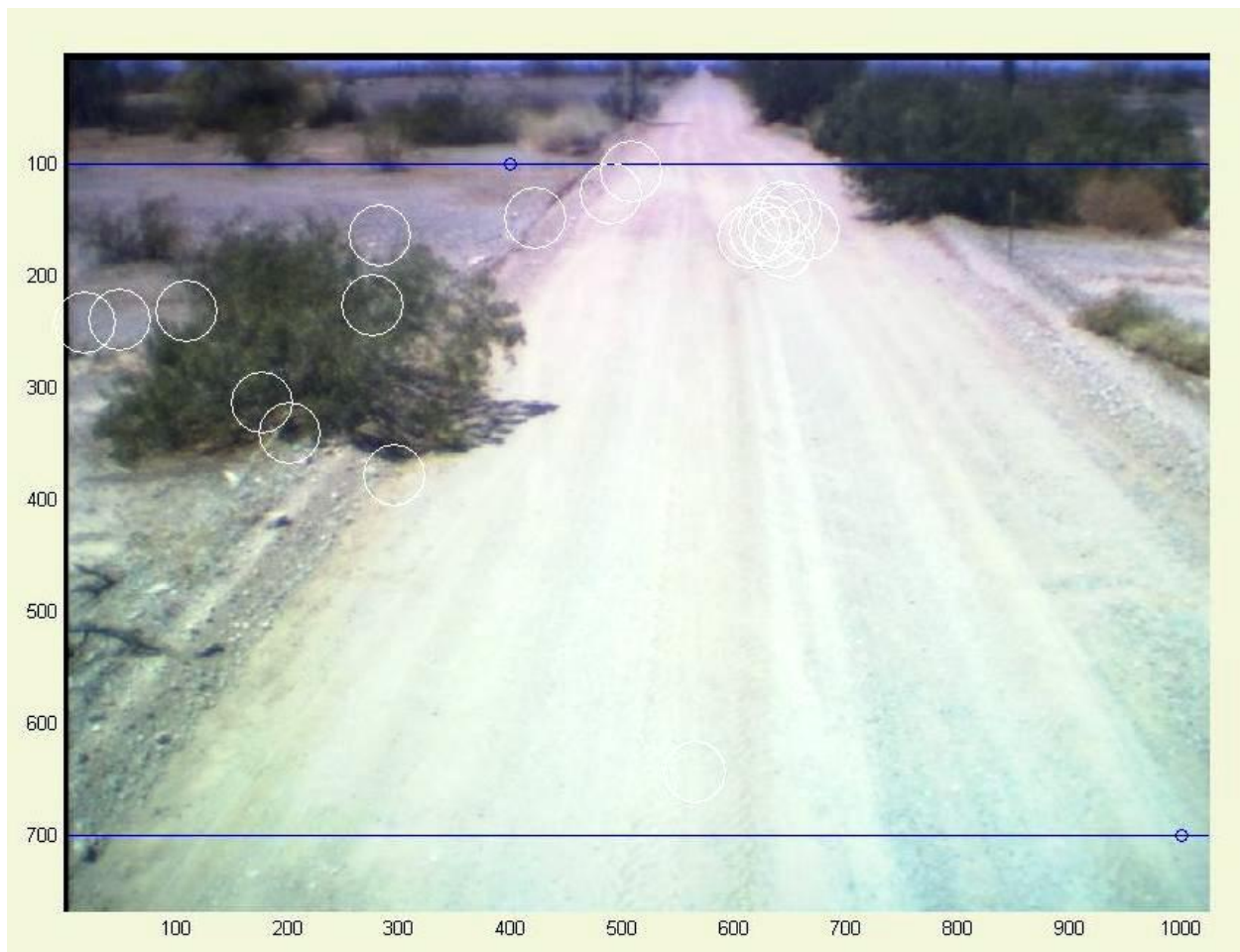


Figure 31. One frame of a typical video image sequence. The faint white circles indicate potential target hits in this video frame.

Technical Significance and Relevance to Army

The Army needs to detect landmines and more generally explosive devices at greater standoff distances. Ground Penetrating Radar has been shown to give excellent results in the downward looking scenarios. Recent experiments have demonstrated the potential for landmine and explosive detection with forward looking GPR. This project aims at investigation of salient features for discriminating between IEDs and clutter objects in FLGPR and corresponding investigations in color and FLIR imagery. In particular, exploiting features present in images can significantly reduce the number of false alarms found in FLGPR detection algorithms. The results of this project will facilitate the utilization of FLGPR in the next generation of vehicle mounted explosive detection systems. The inclusion of Forwarding-Looking IR (FLIR) and color for IED detection shows promise. Our particular approach deals with increasing our understanding of the interaction between FLGPR and EO imagery.

While completely automated algorithms are a laudable goal, the continually evolving nature of the explosive hazard threat has caused the Army to reexamine available approaches. The brain of a trained human operator is a superior object recognition “machine” when not overloaded by massive amounts of data. In this project, we are studying the fusion of features and algorithms derived from various streams of sensor imagery, specifically color and various infra-red ranges for

cuing an operator to likely places to search for explosive devices. Since we can map these image streams onto FLGPR coordinates, results from this part of the project can be combine with FLGPR for increased detection capabilities, further relieving the human from the tedious task of searching large amounts of uninteresting data and enabling him or her to concentrate on the infrequent, but important parts of the scene.

Recent Accomplishments

- Continued development of fusion algorithms for camera imagery and FLGPR array data;
 - Achieved 0.05 FA/m² at 90% POD on preliminary Army test lanes.
- Continued development of video sequence to UTM coordinate transformation;
 - code is being developed to distribute to multi-university and government team.
- Developed an image feature library extraction suite of algorithms to assist in building feature sets for training of classifiers and fusion of multiple modalities;
 - Used to study classes of clutter and code distributed on an restricted website for cleared research participants.
- Researched multiple instance learning on above ground targets in color image sequences;
 - Improved learning classifier parameters (better matching to targets as they appear in video frames).
- Investigated spatial spectrum features on the complex FLGPR array data;
 - Considerably increased POD with decreased FAR on Army test lane data over standard magnitude feature.
- Investigated fusion of LWIR and Color imagery in change detection scenario;
 - much lower FAR at constant POD compared to direct detection.

Technology Transfer

We are in close contact with several appropriate personnel at RDECOM CERDEC NVESD. All algorithms, code, documentation, and results are regularly transferred to them. We have posted several code modules on a restricted website hosted by the University of Florida to facilitate collaboration among algorithm developers and the Government. We held discussions with NVESD personnel during the SPIE meeting on directions of Forward Looking Explosive Hazard detection. Robert Luke, Keller's PhD student, took a position in the Countermine Division of NVESD upon completion of his PhD at MU.

Students Funded

Timothy Havens, PhD – graduated August 2010

Kevin Stone, MS

Chris Spain, MS

Justin Farrell, MS

Bradley Calhoun, BS

Donald Schartman, BS

Additionally, Dr. Mihail Popescu (MU), Dr. Garrison Greenwood (Portland State University) and Scott Blakely (Portland State University) worked on this project.

Publications

Because the work on this project involves explosive devices, we had to be careful with respect to publication. However, the following publications relate to aspects of this project.

Greenwood, G., Blakely, S., Schartman, D., Calhoun, B., Keller, J., Ton, T. Wong, D., and Soumekh, M., “Feature Extraction and Object Recognition in Multi-Modal Forward Looking Imagery”, *Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV*, Orlando, FL, April, 2010.

Havens, T., Ho, K.C., Farrell, J., Keller, J., Popescu, M., Ton, T. Wong, D., and Soumekh, M., “Locally-Adaptive Detection Algorithm for Forward-Looking Ground-Penetrating Radar”, *Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV*, Orlando, FL, April, 2010.

Popescu, M., Stone, K., Havens, T., Ho, K.C., and Keller, J., “Anomaly Detection in Forward Looking Infrared Imaging Using One Class Classifiers”, *Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV*, Orlando, FL, April, 2010.

Stone, K., Keller, J., Popescu, M., Havens, T., and Ho, K.C., Forward Looking Anomaly Detection via Fusion of Infrared and Color Imagery, *Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV*, Orlando, FL, April, 2010.

Havens, T., Spain, C., Ho, K.C., Keller, J., Ton, T. Wong, D., and Soumekh, M., “Improved Detection and False Alarm Rejection Using FLGPR and Color Imagery in a Forward-Looking System”, *Proceedings, SPIE Conference on Detection And Remediation Technologies For Mines And Minelike Targets XV*, Orlando, FL, April, 2010.

Awards/Honors

- Jim won the 2010 IEEE Computational Intelligence Society Meritorious Service Award
- Jim won the 2010 University of Missouri College of Engineering Senior Faculty Research Award